B6

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072830 A2

(51) International Patent Classification⁷: C12N 15/12, C07K 14/47, G01N 33/53, A61K 39/00

(21) International Application Number: PCT/US02/03715

(22) International Filing Date: 8 February 2002 (08.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/268,111	9 February 2001 (09.02.2001)	US
60/271,175	23 February 2001 (23.02.2001)	US
60/274,503	8 March 2001 (08.03.2001)	US
60/274,552	9 March 2001 (09.03.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Poltan Place Way, San Jose, CA 95121 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **DING, Li** [US/US]; 3353 Alma Street, #146, Palo Alto, CA 94506 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road, #107, Redwood City, CA 94062 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **KHARE, Reena** [CN/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/072830 A2

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: The invention provides human proteins associated with cell growth, differentiation, and death (CGDD) and polynucleotides which identify and encode CGDD. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. THe invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CGDD.

PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, reproductive disorders, and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death.

BACKGROUND OF THE INVENTION

Human growth and development requires the spatial and temporal regulation of cell differentiation, cell proliferation, and apoptosis. These processes coordinately control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. At the cellular level, growth and development is governed by the cell's decision to enter into or exit from the cell division cycle and by the cell's commitment to a terminally differentiated state. These decisions are made by the cell in response to extracellular signals and other environmental cues it receives. The following discussion focuses on the molecular mechanisms of cell division, embryogenesis, cell differentiation and proliferation, and apoptosis, as well as disease states such as cancer which can result from disruption of these mechanisms.

Cell Cycle

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms. In multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Progression through the cell cycle is governed by the intricate interactions of protein complexes. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including cyclins, cyclin-dependent protein kinases, growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, and tumor-suppressor proteins.

Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA,

and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, centromere-associated proteins such as CENP-A, -B, and -C, play structural roles in kinetochore formation and assembly (Saffery, R. et al. (2000) Human Mol. Gen. 9: 175-185).

During the M phase of eukaryotic cell cycling, structural rearrangements occur ensuring appropriate distribution of cellular components between daughter cells. Breakdown of interphase structures into smaller subunits is common. The nuclear envelope breaks into vesicles, and nuclear lamins are disassembled. Subsequent phosphorylation of these lamins occurs and is maintained until telophase, at which time the nuclear lamina structure is reformed. cDNAs responsible for encoding M phase phosphorylation (MPPs) are components of U3 small nucleolar ribonucleoprotein (snoRNP), and relocalize to the nucleolus once mitosis is complete (Westendorf, J.M. et al. (1998) J. Biol. Chem. 9:437-449). U3 snoRNPs are essential mediators of RNA processing events.

Proteins involved in the regulation of cellular processes such as mitosis include the Ser/Thr-protein phosphatases type 1 (PP-1). PP-1s act by dephosphorylation of key proteins involved in the metaphase-anaphase transition. The gene PP1R7 encodes the regulatory polypeptide sds22, having at least six splice variants (Ceulemans, H. et al. (1999) Eur. J. Biochem. 262:36-42). Sds22 modulates the activity of the catalytic subunit of PP-1s, and enhances the PP-1-dependent dephosphorylation of mitotic substrates.

Cell cycle regulatory proteins play an important role in cell proliferation and cancer. For example, failures in the proper execution and timing of cell cycle events can lead to chromosome segregation defects resulting in aneuploidy or polyploidy. This genomic instability is characteristic of transformed cells (Luca, F.C. and Winey, M. (1998) Mol. Biol. Cell. 9:29-46). A recently identified

protein, mMOB1, is the mammalian homolog of yeast MOB1, an essential yeast gene required for completion of mitosis and maintenance of ploidy. The mammalian mMOB1 is a member of protein complexes including protein phosphatase 2A (PP2A), and its phosphorylation appears to be regulated by PP2A (Moreno, C.S. et al. (2001) J. Biol. Chem. 276:24253-24260). PP2A has been implicated in the development of human cancers, including lung and colon cancers and leukemias.

Cell cycle regulation involves numerous proteins interacting in a sequential manner. The eukaryotic cell cycle consists of several highly controlled events whose precise order ensures successful DNA replication and cell division. Cells maintain the order of these events by making later events dependent on the successful completion of earlier events. This dependency is enforced by cellular mechanisms called checkpoints. Examples of additional cell cycle regulatory proteins include the histone deacetylases (HDACs). HDACs are involved in cell cycle regulation, and modulate chromatin structure. Human HDAC1 has been found to interact *in vitro* with the human Hus1 gene product, whose *Schizosaccharomyces pombe* homolog has been implicated in G_2/M checkpoint control (Cai, R.L. et al. (2000) J. Biol. Chem. 275:27909-27916).

DNA damage (G_2) and DNA replication (S-phase) checkpoints arrest eukaryotic cells at the G_2/M transition. This arrest provides time for DNA repair or DNA replication to occur before entry into mitosis. Thus, the G_2/M checkpoint ensures that mitosis only occurs upon completion of DNA replication and in the absence of chromosomal damage. The Hus1 gene of *Schizosaccharomyces pombe* is a cell cycle checkpoint gene, as are the rad family of genes (e.g., rad1 and rad9) (Volkmer, E. and Karnitz, L.M. (1999) J. Biol. Chem. 274:567-570; Kostrub C.F. et al. (1998) EMBO J. 17:2055-2066). These genes are involved in the mitotic checkpoint, and are induced by either DNA damage or blockage of replication. Induction of DNA damage or replication block leads to loss of function of the Hus1 gene and subsequent cell death. Human homologs have been identified for most of the rad genes, including ATM and ATR, the human homologs of rad3p. Mutations in the ATM gene are correlated with the severe congenital disease ataxia-telagiectasia (Savitsky, K. et al. (1995) Science 268:1749-1753). The human Hus1 protein has been shown to act in a complex with rad1 protein which interacts with rad9, making them central components of a DNA damage-responsive protein complex of human cells (Volkmer, E. and Karnitz, L.M. (1999) J. Biol. Chem. 274:567-570).

The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins act by binding to and activating a group of cyclin-dependent protein kinases (Cdks) which then phosphorylate and activate selected proteins involved in the mitotic process. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and Wolgemuth, D.J. (1993) Development 118:229-40). In addition, cyclins contain a conserved 9 amino acid sequence in the N-terminal region of the molecule called the "destruction box". This sequence is

believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (Hunt, T. (1991) Nature 349:100-101). Several types of cyclins exist (Ciechanover, A. (1994) Cell 79:13-21). Progression through G1 and S phase is driven by the G1 cyclins and their catalytic subunits, including Cdk2–cyclin A, Cdk2–cyclin E, Cdk4–cyclin D and Cdk6–cyclin D. Progression through the G2–M transition is driven by the activation of mitotic CDK–cyclin complexes such as Cdc2–cyclin A, Cdc2–cyclin B1 and Cdc2–cyclin B2 complexes (reviewed in Yang, J. and Kornbluth, S. (1999) Trends in Cell Biology 9:207-210).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaroytic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. The UCS is implicated in the degradation of mitotic cyclin kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra).

The process of ubiquitin conjugation and protein degradation occurs in five principle steps (Jentsch, S. (1992) Annu. Rev. Genet. 26:179-207). First ubiquitin (Ub), a small, heat stable protein is activated by a ubiquitin-activating enzyme (E1) in an ATP dependent reaction which binds the C-terminus of Ub to the thiol group of an internal cysteine residue in E1. Second, activated Ub is transferred to one of several Ub-conjugating enzymes (E2). Different ubiquitin-dependent proteolytic pathways employ structurally similar, but distinct ubiquitin-conjugating enzymes that are associated with recognition subunits which direct them to proteins carrying a particular degradation signal. Third, E2 transfers the Ub molecule through its C-terminal glycine to a member of the ubiquitin-protein ligase family, E3. Fourth, E3 transfers the Ub molecule to the target protein. Additional Ub molecules may be added to the target protein forming a multi-Ub chain structure. Fifth, the ubiquinated protein is then recognized and degraded by the proteasome, a large, multisubunit proteolytic enzyme complex, and Ub is released for re-utilization.

Prior to activation, Ub is usually expressed as a fusion protein composed of an N-terminal ubiquitin and a C-terminal extension protein (CEP) or as a polyubiquitin protein with Ub monomers attached head to tail. CEPs have characteristics of a variety of regulatory proteins; most are highly basic, contain up to 30% lysine and arginine residues, and have nucleic acid-binding domains (Monia, B.P. et al. (1989) J. Biol. Chem. 264:4093-4103). The fusion protein is an important intermediate which appears to mediate co-regulation of the cell's translational and protein degradation activities, as well as localization of the inactive enzyme to specific cellular sites. Once delivered, C-terminal hydrolases cleave the fusion protein to release a functional Ub (Monia et al., supra).

Ub-conjugating enzymes (E2s) are important for substrate specificity in different UCS pathways. All E2s have a conserved domain of approximately 16 kDa called the UBC domain that is at least 35% identical in all E2s and contains a centrally located cysteine residue required for ubiquitin-enzyme thiolester formation (Jentsch, supra). A well conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations beyond this conserved domain are used to classify the E2 enzymes. Class I E2s consist almost exclusively of the conserved UBC domain. Class II E2s have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. Class III E2s have unique N-terminal extensions which are believed to be involved in enzyme regulation or substrate specificity.

A mitotic cyclin-specific E2 (E2-C) is characterized by the conserved UBC domain, an N-terminal extension of 30 amino acids not found in other E2s, and a 7 amino acid unique sequence adjacent to this extension. These characteristics together with the high affinity of E2-C for cyclin identify it as a new class of E2 (Aristarkhov, A. et al. (1996) Proc. Natl. Acad. Sci. 93:4294-99).

Ubiquitin-protein ligases (E3s) catalyze the last step in the ubiquitin conjugation process, covalent attachment of ubiquitin to the substrate. E3 plays a key role in determining the specificity of the process. Only a few E3s have been identified so far. One type of E3 ligases is the HECT (homologous to E6-AP C-terminus) domain protein family. One member of the family, E6-AP (E6-associated protein) is required, along with the human papillomavirus (HPV) E6 oncoprotein, for the ubiquitination and degradation of p53 (Scheffner et al. (1993) Cell 75:495-505). The C-terminal domain of HECT proteins contains the highly conserved ubiquitin-binding cysteine residue. The N-terminal region of the various HECT proteins is variable and is believed to be involved in specific substrate recognition (Huibregtse, J.M. et al. (1997) Proc. Natl Acad. Sci. USA 94:3656-3661). The SCF (Skp1-Cdc53/Cullin-F box receptor) family of proteins comprise another group of ubiquitin ligases (Deshaies, R. (1999) Annu. Rev. Dev. Biol. 15:435-467). Multiple proteins are recruited into the SCF complex, including Skp1, cullin, and an F box domain containing protein. The F box protein binds the substrate for the ubiquitination reaction and may play roles in determining substrate specificity and orienting the substrate for reaction. Skp1 interacts with both the F box protein and cullin and may be involved in positioning the F box protein and cullin in the complex for transfer of ubiquitin from the E2 enzyme to the protein substrate. Substrates of SCF ligases include proteins involved in regulation of CDK activity, activation of transcription, signal transduction, assembly of kinetochores, and DNA replication.

Sgt1 was identified in a screen for genes in yeast that suppress defects in kinetochore function caused by mutations in Skp1 (Kitagawa, K. et al. (1999) Mol. Cell 4:21-33). Sgt1 interacts with Skp1 and associates with SCF ubiquitin ligase. Defects in Sgt1 cause arrest of cells at either G1 or G2 stages of the cell cycle. A yeast Sgt1 null mutant can be rescued by human Sgt1, an indication

of the conservation of Sgt1 function across species. Sgt1 is required for assembly of kinetochore complexes in yeast.

Abnormal activities of the UCS are implicated in a number of diseases and disorders. These include, e.g., cachexia (Llovera, M. et al. (1995) Int. J. Cancer 61: 138-141), degradation of the tumor-suppressor protein, p53 (Ciechanover, supra), and neurodegeneration such as observed in Alzheimer's disease (Gregori, L. et al. (1994) Biochem. Biophys. Res. Commun. 203: 1731-1738). Since ubiquitin conjugation is a rate-limiting step in antigen presentation, the ubiquitin degradation pathway may also have a critical role in the immune response (Grant E.P. et al. (1995) J. Immunol. 155: 3750-3758).

Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) BioEssays 17:471-480).

Embryogenesis

Mammalian embryogenesis is a process which encompasses the first few weeks of development following conception. During this period, embryogenesis proceeds from a single fertilized egg to the formation of the three embryonic tissues, then to an embryo which has most of its internal organs and all of its external features.

The normal course of mammalian embryogenesis depends on the correct temporal and spatial regulation of a large number of genes and tissues. These regulation processes have been intensely studied in mouse. An essential process that is still poorly understood is the activation of the embryonic genome after fertilization. As mouse oocytes grow, they accumulate transcripts that are either translated directly into proteins or stored for later activation by regulated polyadenylation. During subsequent meiotic maturation and ovulation, the maternal genome is transcriptionally inert, and most maternal transcripts are deadenylated and/or degraded prior to, or together with, the activation of the zygotic genes at the two-cell stage (Stutz, A. et al. (1998) Genes Dev. 12:2535-2548). The maternal to embryonic transition involves the degradation of oocyte, but not zygotic transcripts, the activation of the embryonic genome, and the induction of cell cycle progression to accommodate early development.

MATER (Maternal Antigen That Embryos Require) was initially identified as a target of antibodies from mice with ovarian immunity (Tong, Z-B., and Nelson, L.M. (1999) Endocrinology 140:3720-3726). Expression of the gene encoding MATER is restricted to the oocyte, making it one of a limited number of known maternal-effect genes in mammals (Tong, Z-B., et al. (2000) Mamm. Genome 11:281-287). The MATER protein is required for embryonic development beyond two cells,

based upon preliminary results from mice in which this gene has been inactivated. The 1111-amino acid MATER protein contains a hydrophilic repeat region in the amino terminus, and a region containing 14 leucine-rich repeats in the carboxyl terminus. These repeats resemble the sequence found in porcine ribonuclease inhibitor that is critical for protein-protein interactions.

The degradation of maternal transcripts during meiotic maturation and ovulation may involve the activation of a ribonuclease just prior to ovulation. Thus the function of MATER may be to bind to the maternal ribonuclease and prevent degradation of zygotic transcripts (Tong (2000) supra). In addition to its role in oocyte development and embryogenesis, MATER may also be relevant to the pathogenesis of ovarian immunity, as it is a target of autoantibodies in mice with autoimmune oophoritis (Tong (1999) supra).

The maternal mRNA D7 is a moderately abundant transcript in Xenopus laevis whose expression is highest in, and perhaps restricted to, oogenesis and early embryogenesis. The D7 protein is absent from oocytes and first begins to accumulate during oocyte maturation. Its levels are highest during the first day of embryonic development and then they decrease. The loss of D7 protein affects the maturation process itself, significantly delaying the time course of germinal vesicle breakdown. Thus, D7 is a newly described protein involved in oocyte maturation (Smith R.C., et al. (1988) Genes Dev. 2(10):1296-306.)

Many other genes are involved in subsequent stages of embryogenesis. After fertilization, the oocyte is guided by fimbria at the distal end of each fallopian tube into and through the fallopian tube and thence into the uterus. Changes in the uterine endometrium prepare the tissue to support the implantation and embryonic development of a fertilized ovum. Several stages of division have occurred before the dividing ovum, now a blastocyst with about 100 cells, enters the uterus. Upon reaching the uterus, the developing blastocyst usually remains in the uterine cavity an additional two to four days before implanting in the endometrium, the inner lining of the uterus. Implantation results from the action of trophoblast cells that develop over the surface of the blastocyst. These cells secrete proteolytic enzymes that digest and liquefy the cells of the endometrium. The invasive process is reviewed in Fisher and Damsky (1993; Semin Cell Biol 4:183-188) and Graham and Lala (1992; Biochem Cell Biol 70:867-874). Once implantation has taken place, the trophoblast and other sublying cells proliferate rapidly, forming the placenta and the various membranes of pregnancy. (See Guyton, A.C. (1991) Textbook of Medical Physiology, 8th ed. W.B. Saunders Company, Philadelphia pp. 915-919.)

The placenta has an essential role in protecting and nourishing the developing fetus. In most species the syncytiotrophoblast layer is present on the outside of the placenta at the fetal-maternal interface. This is a continuous structure, one cell deep, formed by the fusion of the constituent trophoblast cells. The syncytiotrophoblast cells play important roles in maternal-fetal exchange, in

tissue remodeling during fetal development, and in protecting the developing fetus from the maternal immune response (Stoye, J.P. and Coffin, J.M. (2000) Nature 403:715-717).

A gene called syncytin is the envelope gene of a human endogenous defective provirus. Syncytin is expressed in high levels in placenta, and more weakly in testis, but is not detected in any other tissues (Mi, S. et al. (2000) Nature 403:785-789). Syncytin expression in the placenta is restricted to the syncytiotrophoblasts. Since retroviral env proteins are often involved in promoting cell fusion events, it was thought that syncytin might be involved in regulating the fusion of trophoblast cells into the syncytiotrophoblast layer. Experiments demonstrated that syncytin can mediate cell fusion in vitro, and that anti-syncytin antibodies can inhibit the fusion of placental cytotrophoblasts (Mi, supra). In addition, a conserved immunosuppressive domain present in retroviral envelope proteins, and found in syncytin at amino acid residues 373-397, might be involved in preventing maternal immune responses against the developing embryo.

Syncytin may also be involved in regulating trophoblast invasiveness by inducing trophoblast fusion and terminal differentiation (Mi, supra). Insufficient trophoblast infiltration of the uterine wall is associated with placental disorders such as preeclampsia, or pregnancy induced hypertension, while uncontrolled trophoblast invasion is observed in choriocarcinoma and other gestational trophoblastic diseases. Thus syncytin function may be involved in these diseases.

Cell Differentiation

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function, despite the fact that each cell is like the others in its hereditary endowment. Cell differentiation is the process by which cells come to differ in their structure and physiological function. The cells of a multicellular organism all arise from mitotic divisions of a single-celled zygote. The zygote is totipotent, meaning that it has the ability to give rise to every type of cell in the adult body. During development the cellular descendants of the zygote lose their totipotency and become determined. Once its prospective fate is achieved, a cell is said to have differentiated. All descendants of this cell will be of the same type.

Human growth and development requires the spatial and temporal regulation of cell differentiation, along with cell proliferation and regulated cell death. These processes coordinate to control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. The processes involved in cell differentiation are also relevant to disease states such as cancer, in which case the factors regulating normal cell differentiation have been altered, allowing the cancerous cells to proliferate in an anaplastic, or undifferentiated, state.

The mechanisms of differentiation involve cell-specific regulation of transcription and translation, so that different genes are selectively expressed at different times in different cells.

Genetic experiments using the fruit fly Drosophila melanogaster have identified regulated cascades of transcription factors which control pattern formation during development and differentiation. These include the homeotic genes, which encode transcription factors containing homeobox motifs. The products of homeotic genes determine how the insect's imaginal discs develop from masses of undifferentiated cells to specific segments containing complex organs. Many genes found to be involved in cell differentiation and development in Drosophila have homologs in mammals. Some human genes have equivalent developmental roles to their Drosophila homologs. The human homolog of the Drosophila eyes absent gene (eya) underlies branchio-oto-renal syndrome, a developmental disorder affecting the ears and kidneys (Abdelhak, S. et al. (1997) Nat. Genet. 15:157-164). The Drosophila slit gene encodes a secreted leucine-rich repeat containing protein expressed by the midline glial cells and required for normal neural development.

At the cellular level, growth and development are governed by the cell's decision to enter into or exit from the cell cycle and by the cell's commitment to a terminally differentiated state. Differential gene expression within cells is triggered in response to extracellular signals and other environmental cues. Such signals include growth factors and other mitogens such as retinoic acid; cell-cell and cell-matrix contacts; and environmental factors such as nutritional signals, toxic substances, and heat shock. Candidate genes that may play a role in differentiation can be identified by altered expression patterns upon induction of cell differentiation in vitro.

The final step in cell differentiation results in a specialization that is characterized by the production of particular proteins, such as contractile proteins in muscle cells, serum proteins in liver cells and globins in red blood cell precursors. The expression of these specialized proteins depends at least in part on cell-specific transcription factors. For example, the homobox-containing transcription factor PAX-6 is essential for early eye determination, specification of ocular tissues, and normal eye development in vertebrates.

In the case of epidermal differentiation, the induction of differentiation-specific genes occurs either together with or following growth arrest and is believed to be linked to the molecular events that control irreversible growth arrest. Irreversible growth arrest is an early event which occurs when cells transit from the basal to the innermost suprabasal layer of the skin and begin expressing squamous-specific genes. These genes include those involved in the formation of the cross-linked envelope, such as transglutaminase I and III, involucrin, loricin, and small proline-rich repeat (SPRR) proteins. The SPRR proteins are 8-10 kDa in molecular mass, rich in proline, glutamine, and cysteine, and contain similar repeating sequence elements. The SPRR proteins may be structural proteins with a strong secondary structure or metal-binding proteins such as metallothioneins. (Jetten, A. M. and Harvat, B. L. (1997) J. Dermatol. 24:711-725; PRINTS Entry PR00021 PRORICH Small proline-rich protein signature.)

The Wnt gene family of secreted signaling molecules is highly conserved throughout eukaryotic cells. Members of the Wnt family are involved in regulating chondrocyte differentiation within the cartilage template. Wnt-5a, Wnt-5b and Wnt-4 genes are expressed in chondrogenic regions of the chicken limb, Wnt-5a being expressed in the perichondrium (mesenchymal cells immediately surrounding the early cartilage template). Wnt-5a misexpression delays the maturation of chondrocytes and the onset of bone collar formation in chicken limb (Hartmann, C. and Tabin, C.J. (2000) Development 127:3141-3159).

Glypicans are a family of cell surface heparan sulfate proteoglycans that play an important role in cellular growth control and differentiation. Cerebroglycan, a heparan sulfate proteoglycan expressed in the nervous system, is involved with the motile behavior of developing neurons (Stipp, C.S. et al. (1994) J. Cell Biol. 124:149-160).

Notch plays an active role in the differentiation of glial cells, and influences the length and organization of neuronal processes (for a review, see Frisen, J. and Lendahl, U. (2001) Bioessays 23:3-7). The Notch receptor signaling pathway is important for morphogenesis and development of many organs and tissues in multicellular species. Drosophila fringe proteins modulate the activation of the Notch signal transduction pathway at the dorsal-ventral boundary of the wing imaginal disc. Mammalian fringe-related family members participate in boundary determination during segmentation (Johnston, S.H. et al. (1997) Development 124:2245-2254).

Recently a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues called the LIM domain (for Lin-11 Isl-1 Mec-3) (Freyd G. et al. (1990) Nature 344:876-879; Baltz R. et al. (1992) Plant Cell 4:1465-1466). In the LIM domain, there are seven conserved cysteine residues and a histidine. The LIM domain binds two zinc ions (Michelsen J.W. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:4404-4408). LIM does not bind DNA, rather it seems to act as an interface for protein-protein interaction.

Apoptosis

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell

shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, especially proteases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

The Bcl-2 family of proteins, as well as other cytoplasmic proteins, are key regulators of apoptosis. There are at least 15 Bcl-2 family members within 3 subfamilies. These proteins have been identified in mammalian cells and in viruses, and each possesses at least one of four Bcl-2 homology domains (BH1 to BH4), which are highly conserved. Bcl-2 family proteins contain the BH1 and BH2 domains, which are found in members of the pro-survival subfamily, while those proteins which are most similar to Bcl-2 have all four conserved domains, enabling inhibition of apoptosis following encounters with a variety of cytotoxic challenges. Members of the pro-survival subfamily include Bcl-2, Bcl-x_L, Bcl-w, Mcl-1, and A1 in mammals; NF-13 (chicken); CED-9 (Caenorhabditis elegans); and viral proteins BHRF1, LMW5-HL, ORF16, KS-Bcl-2, and E1B-19K. The BH3 domain is essential for the function of pro-apoptosis subfamily proteins. The two pro-apoptosis subfamilies, Bax and BH3, include Bax, Bak, and Bok (also called Mtd); and Bik, Blk, Hrk, BNIP3, Bim_L, Bad, Bid, and Egl-1 (C. elegans); respectively. Members of the Bax subfamily contain the BH1, BH2, and BH3 domains, and resemble Bcl-2 rather closely. In contrast, members of the BH3 subfamily have only the 9-16 residue BH3 domain, being otherwise unrelated to any known protein, and only Bik and Blk share sequence similarity. The proteins of the two pro-apoptosis subfamilies may be the antagonists of pro-survival subfamily proteins. This is illustrated in C. elegans where Egl-1, which is required for apoptosis, binds to and acts via CED-9 (for review, see Adams, J.M. and Cory, S. (1998) Science 281:1322-1326).

Heterodimerization between pro-apoptosis and anti-apoptosis subfamily proteins seems to have a titrating effect on the functions of these protein subfamilies, which suggests that relative concentrations of the members of each subfamily may act to regulate apoptosis. Heterodimerization is not required for a pro-survival protein; however, it is essential in the BH3 subfamily, and less so in the Bax subfamily.

The Bcl-2 protein has 2 isoforms, alpha and beta, which are formed by alternative splicing. It

forms homodimers and heterodimers with Bax and Bak proteins and the Bcl-X isoform Bcl-x_S. Heterodimerization with Bax requires intact BH1 and BH2 domains, and is necessary for pro-survival activity. The BH4 domain seems to be involved in pro-survival activity as well. Bcl-2 is located within the inner and outer mitochondrial membranes, as well as within the nuclear envelope and endoplasmic reticulum, and is expressed in a variety of tissues. Its involvement in follicular lymphoma (type II chronic lymphatic leukemia) is seen in a chromosomal translocation T(14;18) (q32;q21) and involves immunoglobulin gene regions.

The Bcl-x protein is a dominant regulator of apoptotic cell death. Alternative splicing results in three isoforms, Bcl-xB, a long isoform, and a short isoform. The long isoform exhibits cell death repressor activity, while the short isoform promotes apoptosis. Bcl-xL forms heterodimers with Bax and Bak, although heterodimerization with Bax does not seem to be necessary for pro-survival (anti-apoptosis) activity. Bcl-xS forms heterodimers with Bcl-2. Bcl-x is found in mitochondrial membranes and the perinuclear envelope. Bcl-xS is expressed at high levels in developing lymphocytes and other cells undergoing a high rate of turnover. Bcl-xL is found in adult brain and in other tissues' long-lived post-mitotic cells. As with Bcl-2, the BH1, BH2, and BH4 domains are involved in pro-survival activity.

The Bcl-w protein is found within the cytoplasm of almost all myeloid cell lines and in numerous tissues, with the highest levels of expression in brain, colon, and salivary gland. This protein is expressed in low levels in testis, liver, heart, stomach, skeletal muscle, and placenta, and a few lymphoid cell lines. Bcl-w contains the BH1, BH2, and BH4 domains, all of which are needed for its cell survival promotion activity. Although mice in which Bcl-w gene function was disrupted by homologous recombination were viable, healthy, and normal in appearance, and adult females had normal reproductive function, the adult males were infertile. In these males, the initial, prepuberty stage of spermatogenesis was largely unaffected and the testes developed normally. However, the seminiferous tubules were disorganized, contained numerous apoptotic cells, and were incapable of producing mature sperm. This mouse model may be applicable to some cases of human male sterility and suggests that alteration of programmed cell death in the testes may be useful in modulating fertility (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431).

Studies in rat ischemic brain found Bcl-w to be overexpressed relative to its normal low constitutive level of expression in nonischemic brain. Furthermore, in vitro studies to examine the mechanism of action of Bcl-w revealed that isolated rat brain mitochondria were unable to respond to an addition of recombinant Bax or high concentrations of calcium when Bcl-w was also present. The normal response would be the release of cytochrome c from the mitochondria. Additionally, recombinant Bcl-w protein was found to inhibit calcium-induced loss of mitochondrial transmembrane potential, which is indicative of permeability transition. Together these findings

suggest that Bcl-w may be a neuro-protectant against ischemic neuronal death and may achieve this protection via the mitochondrial death-regulatory pathway (Yan, C. et al. (2000) J. Cereb. Blood Flow Metab. 20:620-630).

The bfl-1 gene is an additional member of the Bcl-2 family, and is also a suppressor of apoptosis. The Bfl-1 protein has 175 amino acids, and contains the BH1, BH2, and BH3 conserved domains found in Bcl-2 family members. It also contains a Gln-rich NH2-terminal region and lacks an NH domain 1, unlike other Bcl-2 family members. The mouse A1 protein shares high sequence homology with Bfl-1 and has the 3 conserved domains found in Bfl-1. Apoptosis induced by the p53 tumor suppressor protein is suppressed by Bfl-1, similar to the action of Bcl-2, Bcl-xL, and EBV-BHRF1 (D'Sa-Eipper, C. et al. (1996) Cancer Res. 56:3879-3882). Bfl-1 is found intracellularly, with the highest expression in the hematopoietic compartment, i.e. blood, spleen, and bone marrow; moderate expression in lung, small intestine, and testis; and minimal expression in other tissues. It is also found in vascular smooth muscle cells and hematopoietic malignancies. A correlation has been noted between the expression level of bfl-1 and the development of stomach cancer, suggesting that the Bfl-1 protein is involved in the development of stomach cancer, either in the promotion of cancerous cell survival or in cancer (Choi, S.S. et al. (1995) Oncogene 11:1693-1698).

Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death. Strategies for treatment may involve either reestablishing control over cell cycle progression, or selectively stimulating apoptosis in cancerous cells (Nigg, E.A. (1995) BioEssays 17:471-480). Immunological defenses against cancer include induction of apoptosis in mutant cells by tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Response to mitogenic stresses is frequently controlled at the level of transcription and is coordinated by various transcription factors. For example, the Rel/NF-kappa B family of vertebrate transcription factors plays a pivotal role in inflammatory and immune responses to radiation. The NF-kappa B family includes p50, p52, RelA, RelB, cRel, and other DNA-binding proteins. The p52 protein induces apoptosis, upregulates the transcription factor c-Jun, and activates c-Jun N-terminal kinase 1 (JNK1) (Sun, L. et al. (1998) Gene 208:157-166). Most NF-kappa B proteins form DNA-binding homodimers or heterodimers. Dimerization of many transcription factors is mediated by a conserved sequence known as the bZIP domain, characterized by a basic region followed by a leucine zipper.

The Fas/Apo-1 receptor (FAS) is a member of the tumor necrosis factor (TNF) receptor family. Upon binding its ligand (Fas ligand), the membrane-spanning FAS induces apoptosis by recruiting several cytoplasmic proteins that transmit the death signal. One such protein, termed FAS-associated protein factor 1 (FAF1), was isolated from mice, and it was demonstrated that expression of FAF1 in L cells potentiated FAS-induced apoptosis (Chu, K. et al. (1995) Proc. Natl. Acad. Sci.

USA 92:11894-11898). Subsequently, FAS-associated factors have been isolated from numerous other species, including fruit fly and quail (Frohlich, T. et al. (1998) J. Cell Sci. 111:2353-2363). Another cytoplasmic protein that functions in the transmittal of the death signal from Fas is the Fas-associated death domain protein, also known as FADD. FADD transmits the death signal in both FAS-mediated and TNF receptor-mediated apoptotic pathways by activating caspase-8 (Bang, S. et al. (2000) J. Biol. Chem. 275:36217-36222).

Fragmentation of chromosomal DNA is one of the hallmarks of apoptosis. DNA fragmentation factor (DFF) is a protein composed of two subunits, a 40-kDa caspase-activated nuclease termed DFF40/CAD, and its 45-kDa inhibitor DFF45/ICAD. Two mouse homologs of DFF45/ICAD, termed CIDE-A and CIDE-B, have recently been described (Inohara, N. et al. (1998) EMBO J. 17:2526-2533). CIDE-A and CIDE-B expression in mammalian cells activated apoptosis, while expression of CIDE-A alone induced DNA fragmentation. In addition, FAS-mediated apoptosis was enhanced by CIDE-A and CIDE-B, further implicating these proteins as effectors that mediate apoptosis.

Transcription factors play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, are involved in the initiation and execution phases of apoptosis. The activation of the caspases results from the competitive action of the pro-survival and pro-apoptosis Bcl-2-related proteins (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431). A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues.

Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein-associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer.

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells.

(Reviewed in Nagata, S. (1997) Cell 88:355-365.) Binding of TNF to its receptor triggers a signal transduction pathway that results in the activation of a proteolytic caspase cascade. One such caspase, ICE (Interleukin-1β converting enzyme), is a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage (Dinarello, C. A. (1994) FASEB J. 8:1314-1325). ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1β, into its active form, which plays a central role in acute and chronic inflammation, bone resorption, myelogenous leukemia, and other pathological processes. ICE and related caspases cause apoptosis when overexpressed in transfected cell lines.

A caspase recruitment domain (CARD) is found within the prodomain of several apical caspases and is conserved in several apoptosis regulatory molecules such as Apaf-2, RAIDD, and cellular inhibitors of apoptosis proteins (IAPs) (Hofmann, K. et al. (1997) Trends Biochem. Sci. 22:155-157). The regulatory role of CARD in apoptosis may be to allow proteins such as Apaf-1 to associate with caspase-9 (Li, P. et al. (1997) Cell 91:479-489). A human cDNA encoding an apoptosis repressor with a CARD (ARC) which is expressed in both skeletal and cardiac muscle has been identified and characterized. ARC functions as an inhibitor of apoptosis and interacts selectively with caspases (Koseki, T. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5156-5160). All of these interactions have clear effects on the control of apoptosis (reviewed in Chan S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

ES18 was identified as a potential regulator of apoptosis in mouse T-cells (Park, E.J. et al. (1999) Nuc. Acid. Res. 27:1524-1530). ES18 is 428 amino acids in length, contains an N-terminal proline-rich region, an acidic glutamic acid-rich domain, and a putative LXXLL nuclear receptor binding motif. The protein is preferentially expressed in lymph nodes and thymus. The level of ES18 expression increases in T-cell thymoma S49.1 in response to treatment with dexamethasone, staurosporine, or C2-ceramide, which induce apoptosis. ES18 may play a role in stimulating apoptotic cell death in T-cells.

The rat ventral prostate (RVP) is a model system for the study of hormone-regulated apoptosis. RVP epithelial cells undergo apoptosis in response to androgen deprivation. Messenger RNA (mRNA) transcripts that are up-regulated in the apoptotic RVP have been identified (Briehl, M. M. and Miesfeld, R. L. (1991) Mol. Endocrinol. 5:1381-1388). One such transcript encodes RVP.1, the precise role of which in apoptosis has not been determined. The human homolog of RVP.1, hRVP1, is 89% identical to the rat protein (Katahira, J. et al. (1997) J. Biol. Chem. 272:26652-26658). hRVP1 is 220 amino acids in length and contains four transmembrane domains. hRVP1 is highly expressed in the lung, intestine, and liver. Interestingly, hRVP1 functions as a low affinity receptor for the Clostridium perfringens enterotoxin, a causative agent of diarrhea in humans and

other animals.

Cytokine-mediated apoptosis plays an important role in hematopoiesis and the immune response. Myeloid cells, which are the stem cell progenitors of macrophages, neutrophils, erythrocytes, and other blood cells, proliferate in response to specific cytokines such as granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin-3 (IL-3). When deprived of GM-CSF or IL-3, myeloid cells undergo apoptosis. The murine *requiem* (*req*) gene encodes a putative transcription factor required for this apoptotic response in the myeloid cell line FDCP-1 (Gabig, T. G. et al. (1994) J. Biol. Chem. 269:29515-29519). The Req protein is 371 amino acids in length and contains a nuclear localization signal, a single *Kruppel*-type zinc finger, an acidic domain, and a cluster of four unique zinc-finger motifs enriched in cysteine and histidine residues involved in metal binding. Expression of *req* is not myeloid- or apoptosis-specific, suggesting that additional factors regulate Req activity in myeloid cell apoptosis.

Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast, excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) Science 267:1456-1462.)

Impaired regulation of apoptosis is also associated with loss of neurons in Alzheimer's disease. Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. B-amyloid peptide participates in signaling pathways that induce apoptosis and lead to the death of neurons (Kajkowski, C. et al. (2001) J. Biol. Chem. 276:18748-18756). Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

Cancer

Cancer remains a major public health cancer, and current preventative measures and treatments do not match the needs of most patients. Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death. Understanding of the neoplastic process can be aided by the identification of molecular

markers of prognostic and diagnostic importance. Cancers are associated with oncoproteins which are capable of transforming normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein while others are abnormally expressed with respect to location or level of expression. Normal cell proliferation begins with binding of a growth factor to its receptor on the cell membrane, resulting in activation of a signal system that induces and activates nuclear regulatory factors to initiate DNA transcription, subsequently leading to cell division. Classes of oncoproteins known to affect the cell cycle controls include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Several types of cancer-specific genetic markers, such as tumor antigens and tumor suppressors, have also been identified.

Oncogenes

Oncoproteins are encoded by genes, called oncogenes, that are derived from genes that normally control cell growth and development. Many oncogenes have been identified and characterized. These include growth factors such as *sis*, receptors such as *erbA*, *erbB*, *neu*, and *ros*, intracellular receptors such as *src, yes, fps, abl*, and *met*, protein-serine/threonine kinases such as *mos* and *raf*, nuclear transcription factors such as *jun, fos, myc*, *N-myc, myb, ski, and rel*, cell cycle control proteins such as *RB* and *p53*, mutated tumor-suppressor genes such as *mdm2, Cip1, p16*, and *cyclin D, ras*, *set*, *can, sec*, and *gag R10*.

Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22. The hybrid c-abl-bcr gene encodes a chimeric protein that has tyrosine kinase activity. In chronic myeloid leukemia, the chimeric protein has a molecular weight of 210 kd, whereas in acute leukemias a more active 180 kd tyrosine kinase is formed (Robbins, S.L. et al. (1994) Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA).

The Ras superfamily of small GTPases is involved in the regulation of a wide range of cellular signaling pathways. Ras family proteins are membrane-associated proteins acting as molecular switches that bind GTP and GDP, hydrolyzing GTP to GDP. The GTPase-activating protein of Ras (RasGAP) is activated by the GTPase-activating family of proteins (GAPs). A central conserved GAP-related domain, and a C-terminal pleckstrin homology (PH) domain are characteristic of the GAP1 subfamily of RasGAP proteins (Allen, M. et al., (1998) Gene 218:17-25). In the active GTP-bound state Ras family proteins interact with a variety of cellular targets to activate downstream

signaling pathways. For example, members of the Ras subfamily are essential in transducing signals from receptor tyrosine kinases (RTKs) to a series of serine/threonine kinases which control cell growth and differentiation. Activated Ras genes were initially found in human cancers and subsequent studies confirmed that Ras function is critical in the determination of whether cells continue to grow or become terminally differentiated. Stimulation of cell surface receptors activates Ras which, in turn, activates cytoplasmic kinases. The kinases translocate to the nucleus and activate key transcription factors that control gene expression and protein synthesis (Barbacid, M. (1987) Annu. Rev. Biochem. 56:779-827, Treisman, R. (1994) Curr. Opin. Genet. Dev. 4:96-98). Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated, and cause continuous cell proliferation or cancer.

Activation of Ras family proteins is catalyzed by guanine nucleotide exchange factors (GEFs) which catalyze the dissociation of bound GDP and subsequent binding of GTP. A recently discovered RalGEF-like protein, RGL3, interacts with both Ras and the related protein Rit. Constitutively active Rit, like Ras, can induce oncogenic transformation, although since Rit fails to interact with most known Ras effector proteins, novel cellular targets may be involved in Rit transforming activity. RGL3 interacts with both Ras and Rit, and thus may act as a downstream effector for these proteins (Shao, H. and Andres, D.A. (2000) J. Biol. Chem. 275:26914-26924).

<u>Tumor antigens</u>

Tumor antigens are cell surface molecules that are differentially expressed in tumor cells relative to non-tumor tissues. Tumor antigens make tumor cells immunologically distinct from normal cells and are potential diagnostics for human cancers. Several monoclonal antibodies have been identified which react specifically with cancerous cells such as T-cell acute lymphoblastic leukemia and neuroblastoma (Minegishi et al. (1989) Leukemia Res. 13:43-51; Takagi et al. (1995) Int. J. Cancer 61:706-715). In addition, the discovery of high level expression of the HER2 gene in breast tumors has led to the development of therapeutic treatments (Liu et al. (1992) Oncogene 7: 1027-1032; Kern (1993) Am. J. Respir. Cell Mol. Biol. 9:448-454). Tumor antigens are found on the cell surface and have been characterized either as membrane proteins or glycoproteins. For example, MAGE genes encode a family of tumor antigens recognized on melanoma cell surfaces by autologous cytolytic T lymphocytes. Among the 12 human MAGE genes isolated, half are differentially expressed in tumors of various histological types (De Plaen et al. (1994) Immunogenetics 40:360-369). None of the 12 MAGE genes, however, is expressed in healthy tissues except testis and placenta.

TA1, a tumor-associated gene, was identified and cloned based on its increased expression in rat hepatoma cells compared to normal rat liver (Sang, J. et al. (1995) Cancer Res. 55:1152-1159). The deduced amino acid sequence encodes an integral membrane protein which contains multiple

transmembrane domains. TA1 exhibits an oncofetal expression pattern in liver. Transcripts for TA1 are present in rat fetal liver and hepatoma, but they are not present in normal adult rat liver. In normal adult rat, TA1 is expressed at moderate-to-high levels in testes and brain, and at low levels in ovary, spleen, mammary gland, and uterus. TA1 expression is most abundant in placenta, which suggests a developmental role for the molecule (Sang et al., supra).

The E16 gene cloned from human peripheral blood lymphocytes encodes a 241 amino acid integral membrane protein with multiple predicted transmembrane domains (Gaugitsch, H.W. et al. (1992) J. Biol. Chem. 267:11267-73). E16 gene expression is closely linked to cellular activation and division. In myeloid and lymphoid cells, E16 transcripts are rapidly induced and rapidly degraded after stimulation. This pattern of expression resembles the kinetics seen for proto-oncogenes and lymphokines in the T cell system (Gaugitsch et al., supra). E16 expression was not detected in normal (non-cancerous) human tissues such as adult brain, lung, liver, colon, esophagus, stomach, or kidney, nor in four-month fetal brain, lung, liver, or kidney (Wolf, D.A. et al. (1996) Cancer Res. 56:5012-5022; Gaugitsch et al., supra). E16 was detected in every cell line tested (Gaugitsch et al., supra). Its presence in rapidly dividing cell lines and its absence in human tissues with low proliferative potential suggest a direct involvement of E16 protein in the cell division process (Gaugitsch et al., supra).

The proteins encoded by the rat TA1 and human E16 genes share 95% amino acid sequence identity (Wolf et al., supra). Nucleotide probes and antibodies specific for homologous regions of TA1 and E16 were prepared in order to detect TA1/E16 expression in various human cancers. With these probes, elevated levels of TA1/E16 expression were detected in colonic, gastric, and breast adenocarcinomas, and in lymphoma. Although E16 was originally described by Gaugitsch et al. (supra) as a lymphocyte activation marker, no significant levels of TA1/E16 message was detected in tissues from patients with active ulerative colitis and Crohn's disease (Wolf et al., supra).

The TA1 and E16 proteins show significant homology to a putative amino acid permease from the helminth Schistosoma mansoni (GenBank 407047; unpublished). These sequence similarities suggest a potential role for TA1 and E16 proteins in amino acid or nutrient uptake which may be up-regulated in tumor cells (Wolf et al., supra).

Tumor suppressors

Tumor suppressor genes are generally defined as genetic elements whose loss or inactivation contributes to the deregulation of cell proliferation and the pathogenesis and progression of cancer. Tumor suppressor genes normally function to control or inhibit cell growth in response to stress and to limit the proliferative life span of the cell. Several tumor suppressor genes have been identified including the genes encoding the retinoblastoma (Rb) protein, p53, and the breast cancer 1 and 2 proteins (BRCA1 and BRCA2). Mutations in these genes are associated with acquired and inherited

genetic predisposition to the development of certain cancers.

The role of p53 in the pathogenesis of cancer has been extensively studied. (Reviewed in Aggarwal, M. L. et al. (1998) J. Biol. Chem. 273:1-4; Levine, A. (1997) Cell 88:323-331.) About 50% of all human cancers contain mutations in the *p53* gene. These mutations result in either the absence of functional p53 or, more commonly, a defective form of p53 which is overexpressed. p53 is a transcription factor that contains a central core domain required for DNA binding. Most cancer-associated mutations in p53 localize to this domain. In normal proliferating cells, p53 is expressed at low levels and is rapidly degraded. p53 expression and activity is induced in response to DNA damage, abortive mitosis, and other stressful stimuli. In these instances, p53 induces apoptosis or arrests cell growth until the stress is removed. Downstream effectors of p53 activity include apoptosis-specific proteins and cell cycle regulatory proteins, including Rb, oncogene products, cyclins, and cell cycle-dependent kinases.

A novel gene, *ING1*, encoding a potential tumor suppressor protein has been cloned. (Garkavtsev, I. et al. (1996) Nat. Genet. 14:415-420.) Overexpression of *ING1* in normal and transformed cell lines inhibits their growth in vitro. Furthermore, expression of antisense *ING1* promotes neoplastic transformation of cultured cells, as demonstrated by their ability to grow in soft agar and to induce tumors when injected into immunodeficient mice. p33, the protein encoded by *ING1*, localizes to the nucleus and has similarity to retinoblastoma binding protein 2 (RbBP2) and to zinc finger motifs. Decreased expression of p33 is observed in some breast cancer cell lines, and a truncated form of p33 is expressed at high levels in a neuroblastoma cell line. Truncated p33 results from genomic rearrangement at the *ING1* locus. Moreover, levels of *ING1* RNA and protein are increased about 10-fold in senescent cells, which are ageing, non-proliferative cells, compared to the levels expressed in young, proliferating cells. (Garkavtsev, I. and Riabowol, K. (1997) Mol. Cell Biol. 17:2014-2019.) These observations indicate that p33 normally functions to inhibit cell growth and limit cellular life span.

Recent studies show that p33 cooperates with p53 in the negative regulation of cell proliferation. (Garkavtsev, I. et al. (1998) Nature 391:295-298.) The functions of p53 and p33 are interdependent, and p33 directly modulates p53-dependent transcriptional activation. A direct physical association between p33 and p53 has been demonstrated by co-immunoprecipitation, indicating that p33 may influence the activity of p53 in cell cycle control, ageing, and apoptosis.

The metastasis-suppressor gene KAI1 (CD82) has been reported to be related to the tumor suppressor gene p53. KAI1 is involved in the progression of human prostatic cancer and possibly lung and breast cancers when expression is decreased. KAI1 encodes a member of a structurally distinct family of leukocyte surface glycoproteins. The family is known as either the tetraspan transmembrane protein family or transmembrane 4 superfamily (TM4SF) as the members of this

family span the plasma membrane four times. The family is composed of integral membrane proteins having a N-terminal membrane-anchoring domain which functions as both a membrane anchor and a translocation signal during protein biosynthesis. The N-terminal membrane-anchoring domain is not cleaved during biosynthesis. TM4SF proteins have three additional transmembrane regions, seven or more conserved cysteine residues, are similar in size (218 to 284 residues), and all have a large extracellular hydrophilic domain with three potential N-glycosylation sites. The promoter region contains many putative binding motifs for various transcription factors, including five AP2 sites and nine SpI sites. Gene structure comparisons of KAI1 and seven other members of the TM4SF indicate that the splicing sites relative to the different structural domains of the predicted proteins are conserved. This suggests that these genes are related evolutionarily and arose through gene duplication and divergent evolution (Levy, S. et al. (1991) J. Biol. Chem. 266:14597-14602; Dong, J.T. et al. (1995) Science 268:884-886; Dong, J.T. et al., (1997) Genomics 41:25-32).

The Leucine-rich gene-Glioma Inactivated (LGI1) protein shares homology with a number of transmembrane and extracellular proteins which function as receptors and adhesion proteins. LGI1 is encoded by an LLR (leucine-rich, repeat-containing) gene and maps to 10q24. LGI1 has four LLRs which are flanked by cysteine-rich regions and one transmembrane domain (Somerville, R.P., et al. (2000) Mamm. Genome 11:622-627). LGI1 expression is seen predominantly in neural tissues, especially brain. The loss of tumor suppressor activity is seen in the inactivation of the LGI1 protein which occurs during the transition from low to high-grade tumors in malignant gliomas. The reduction of LGI1 expression in low grade brain tumors and its significant reduction or absence of expression in malignant gliomas suggests that it could be used for diagnosis of glial tumor progression (Chernova, O.B., et al. (1998) Oncogene 17:2873-2881).

The ST13 tumor suppressor was identified in a screen for factors related to colorectal carcinomas by subtractive hybridization between cDNA of normal mucosal tissues and mRNA of colorectal carcinoma tissues (Cao, J. et al. (1997) J. Cancer Res. Clin. Oncol. 123:447-451). ST13 is down-regulated in human colorectal carcinomas.

Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene are associated with retinal and central nervous system hemangioblastomas, clear cell renal carcinomas, and pheochromocytomas (Hoffman, M. et al. (2001) Hum. Mol. Genet. 10:1019-1027; Kamada, M. (2001) Cancer Res. 61:4184-4189). Tumor progression is linked to defects or inactivation of the VHL gene. VHL regulates the expression of transforming growth factor-α, the GLUT-1 glucose transporter and vascular endothelial growth factor. The VHL protein associates with elongin B, elongin C, Cul2 and Rbx1 to form a complex that regulates the transcriptional activator hypoxia-inducible factor (HIF). HIF induces genes involved in angiogenesis such as vascular endothelial growth factor and platelet-derived growth factor B. Loss of control of HIF caused by defects in VHL results in the excessive

production of angiogenic peptides. VHL may play roles in inhibition of angiogenesis, cell cycle control, fibronectin matrix assembly, cell adhesion, and proteolysis.

Mutations in tumor suppressor genes are a common feature of many cancers and often appear to affect a critical step in the pathogenesis and progression of tumors. Accordingly, Chang, F. et al. (1995; J. Clin. Oncol. 13: 1009-1022) suggest that it may be possible to use either the gene or an antibody to the expressed protein 1) to screen patients at increased risk for cancer, 2) to aid in diagnosis made by traditional methods, and 3) to assess the prognosis of individual cancer patients. In addition, Hamada, K et al. (1996; Cancer Res. 56:3047-3054) are investigating the introduction of p53 into cervical cancer cells via an adenoviral vector as an experimental therapy for cervical cancer.

The PR-domain genes were recently recognized as playing a role in human tumorigenesis. PR-domain genes normally produce two protein products: the PR-plus product, which contains the PR domain, and the PR-minus product which lacks this domain. In cancer cells, PR-plus is disrupted or overexpressed, while PR-minus is present or overexpressed. The imbalance in the amount of these two proteins appears to be an important cause of malignancy (Jiang, G.L. and Huang, S. (2000) Histol. Histopathol. 15:109-117).

Many neoplastic disorders in humans can be attributed to inappropriate gene transcription. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. An important class of transcriptional regulators are the zinc finger proteins. The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type zinc fingers, and the PHD domain (Lewin, supra; Aasland, R., et al. (1995) Trends Biochem. Sci. 20:56-59). One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

Tumor responsive proteins

Cancers, also called neoplasias, are characterized by continuous and uncontrolled cell proliferation. They can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant growths of soft epithelial cells that may infiltrate surrounding tissues and give rise to metastatic tumors. Sarcomas may be of epithelial origin or arise from connective tissue. Leukemias are progressive malignancies of blood-forming tissue characterized by proliferation of leukocytes and their precursors, and may be classified as myelogenous (granulocyte- or monocyte-

derived) or lymphocytic (lymphocyte-derived). Tumorigenesis refers to the progression of a tumor's growth from its inception. Malignant cells may be quite similar to normal cells within the tissue of origin or may be undifferentiated (anaplastic). Tumor cells may possess few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contains large areas of ischemic necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin. Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or through hematogenous spread. Cancer remains a major public health concern and current preventative measures and treatments do not match the needs of most patients. Understanding of the neoplastic process of tumorigenesis can be aided by the identification of molecular markers of prognostic and diagnostic importance.

Current forms of cancer treatment include the use of immunosuppressive drugs (Morisaki, T. et al. (2000) Anticancer Res. 20: 3363-3373; Geoerger, B. et al. (2001) Cancer Res. 61: 1527-1532). The identification of proteins involved in cell signaling, and specifically proteins that act as receptors for immunosuppressant drugs, may facilitate the development of anti-tumor agents. For example, immunophilins are a family of conserved proteins found in both prokaryotes and eukaryotes that bind to immunosuppressive drugs with varying degrees of specificity. One such group of immunophilic proteins is the peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) family (PPIase, rotamase). These enzymes, first isolated from porcine kidney cortex, accelerate protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (Fischer, G. and Schmid, F.X. (1990) Biochemistry 29: 2205-2212). Included within the immunophilin family are the cyclophilins (e.g., peptidyl-prolyl isomerase A or PPIA) and FK-binding protein (e.g., FKBP) subfamilies. Cyclophilins are multifunctional receptor proteins which participate in signal transduction activities, including those mediated by cyclosporin (or cyclosporine). The PPIase domain of each family is highly conserved between species. Although structurally distinct, these multifunctional receptor proteins are involved in numerous signal transduction pathways, and have been implicated in folding and trafficking events.

The immunophilin protein cyclophilin binds to the immunosuppressant drug cyclosporin A. FKBP, another immunophilin, binds to FK506 (or rapamycin). Rapamycin is an immunosuppressant agent that arrests cells in the G_1 phase of growth, inducing apoptosis. Like cyclophilin, this macrolide antibiotic (produced by Streptomyces tsukubaensis) acts by binding to ubiquitous, predominantly cytosolic immunophilin receptors. These immunophilin/immunosuppressant complexes (e.g., cyclophilin A/cyclosporin A (CypA/CsA) and FKBP12/FK506) achieve their therapeutic results through inhibition of the phosphatase calcineurin, a calcium/calmodulin-dependent protein kinase that participates in T-cell activation (Hamilton, G.S. and Steiner, J.P. (1998) J. Med. Chem. 41: 5119-

5143). The murine fkbp51 gene is abundantly expressed in immunological tissues, including the thymus and T lymphocytes (Baughman, G. et al. (1995) Molec. Cell. Biol. 15: 4395-4402). FKBP12/rapamycin-directed immunosuppression occurs through binding to TOR (yeast) or FRAP (FKBP12-rapamycin-associated protein, in mammalian cells), the kinase target of rapamycin essential for maintaining normal cellular growth patterns. Dysfunctional TOR signaling has been linked to various human disorders including cancer (Metcalfe, S.M. et al. (1997) Oncogene 15: 1635-1642; Emami, S. et al. (2001) FASEB J. 15: 351-361), and autoimmunity (Damoiseaux, J.G. et al. (1996) Transplantation 62: 994-1001).

Several cyclophilin isozymes have been identified, including cyclophilin B, cyclophilin C, mitochondrial matrix cyclophilin, bacterial cytosolic and periplasmic PPIases, and natural-killer cell cyclophilin-related protein possessing a cyclophilin-type PPIase domain, a putative tumor-recognition complex involved in the function of natural killer (NK) cells. These cells participate in the innate cellular immune response by lysing virally-infected cells or transformed cells. NK cells specifically target cells that have lost their expression of major histocompatibility complex (MHC) class I genes (common during tumorigenesis), endowing them with the potential for attenuating tumor growth. A 150-kDa molecule has been identified on the surface of human NK cells that possesses a domain which is highly homologous to cyclophilin/peptidyl-prolyl cis-trans isomerase. This cyclophilin-type protein may be a component of a putative tumor-recognition complex, a NK tumor recognition sequence (NK-TR) (Anderson, S.K. et al. (1993) Proc. Natl. Acad. Sci. USA 90: 542-546). The NKTR tumor recognition sequence mediates recognition between tumor cells and large granular lymphocytes (LGLs), a subpopulation of white blood cells (comprised of activated cytotoxic T cells and natural killer cells) capable of destroying tumor targets. The protein product of the NKTR gene presents on the surface of LGLs and facilitates binding to tumor targets. More recently, a mouse Nktr gene and promoter region have been located on chromosome 9. The gene encodes a NK-cell-specific 150-kDa protein (NK-TR) that is homologous to cyclophilin and other tumor-responsive proteins (Simons-Evelyn, M. et al. (1997) Genomics 40: 94-100).

Other proteins that interact with tumorigenic tissue include cytokines such as tumor necrosis factor (TNF). The TNF family of cytokines are produced by lymphocytes and macrophages, and can cause the lysis of transformed (tumor) endothelial cells. Endothelial protein 1 (Edp1) has been identified as a human gene activated transcriptionally by TNF-alpha in endothelial cells, and a TNF-alpha inducible Edp1 gene has been identified in the mouse (Swift, S. et al. (1998) Biochim. Biophys. Acta 1442: 394-398).

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression

of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new proteins associated with cell growth, differentiation, and death, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders including cancer, developmental disorders, neurological disorders, reproductive disorders, and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteins associated with cell growth, differentiation, and death.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteins associated with cell growth, differentiation, and death, referred to collectively as "CGDD" and individually as "CGDD-1," "CGDD-2," "CGDD-3," "CGDD-4," "CGDD-5," "CGDD-6," "CGDD-7," "CGDD-8," "CGDD-9," "CGDD-10," "CGDD-11," and "CGDD-12." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of

SEQ ID NO:13-24.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the

polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with

decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID

NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CGDD" refers to the amino acid sequences of substantially purified CGDD obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CGDD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with CGDD or by acting on components of the biological pathway in which CGDD participates.

An "allelic variant" is an alternative form of the gene encoding CGDD. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CGDD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CGDD or a polypeptide with at least one functional characteristic of CGDD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CGDD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CGDD. The encoded protein may also be "altered," and may contain deletions, insertions, or

substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CGDD. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CGDD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CGDD. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with CGDD or by acting on components of the biological pathway in which CGDD participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CGDD polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures

on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an <u>in vitro</u> evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed <u>in vivo</u>. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule..

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CGDD, or of any oligopeptide

thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CGDD or fragments of CGDD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr

Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of CGDD or the polynucleotide encoding CGDD which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain

defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent

similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a

standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a

complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate

to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CGDD which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CGDD which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CGDD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CGDD.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CGDD may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in

the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CGDD.

"Probe" refers to nucleic acid sequences encoding CGDD, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the

selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CGDD, nucleic acids encoding CGDD, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the

art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteins associated with cell growth, differentiation, and death (CGDD), the polynucleotides encoding CGDD, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, reproductive disorders, and autoimmune/inflammatory disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteins associated with cell growth, differentiation, and death. For example, SEQ ID NO:3 is 45% identical, from residue M1 to residue I454, to rat RING finger protein terf (GenBank ID g5114353) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.2e-102, which

indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains SPRY, zinc finger (C3HC4 type; RING finger), B-box zinc finger domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a RING finger protein.

In another example, SEQ ID NO:5 is 59% identical, from residue E14 to residue S1159, to human nGAP (GenBank ID g4105589) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a GTPase-activator protein for Ras-like GTPase as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from PROFILESCAN analysis provide further corroborative evidence that SEQ ID NO:5 is a Ras-specific GTPase-activating protein.

In another example, SEQ ID NO:7 is 82% identical, from residue M1 to residue R579, to Rattus norvegicus cerebroglycan (GenBank ID g440127) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.4e-260, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a glypican domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:7 is a glypican.

For example, SEQ ID NO:9 is 99% identical, from residue M1 to residue D448, to the human TRPM-2 gene product (GenBank ID g339973) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.9e-244, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a clusterin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:9 is a clusterin. SEQ ID NO:1-2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 and SEQ ID NO:10-12 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-12 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence

identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_$XXXXXX$_N_1_N_2_$YYYYY$_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis

methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CGDD variants. A preferred CGDD variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CGDD amino acid sequence, and which contains at least one functional or structural characteristic of CGDD.

The invention also encompasses polynucleotides which encode CGDD. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes CGDD. The polynucleotide sequences of SEQ ID NO:13-24, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CGDD. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CGDD. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24 which has at least about 70%, or alternatively at least about 85%, or even at least about

95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CGDD.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CGDD. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CGDD, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding CGDD over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding CGDD. For example, a polynucleotide comprising a sequence of SEQ ID NO:23 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:17 and a polynucleotide comprising a sequence of SEQ ID NO:24 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:21. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CGDD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CGDD, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CGDD, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CGDD and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CGDD under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CGDD or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CGDD and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater

half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CGDD and CGDD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CGDD or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CGDD may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments

adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CGDD may be cloned in recombinant DNA molecules that direct expression of CGDD, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CGDD.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CGDD-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic

oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CGDD, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CGDD may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CGDD itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CGDD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active CGDD, the nucleotide sequences encoding CGDD or

derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CGDD. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CGDD. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CGDD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CGDD and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CGDD. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for

delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CGDD. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CGDD can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CGDD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CGDD are needed, e.g. for the production of antibodies, vectors which direct high level expression of CGDD may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CGDD. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CGDD. Transcription of sequences encoding CGDD may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CGDD may be ligated into

an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CGDD in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CGDD in cell lines is preferred. For example, sequences encoding CGDD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk^- and apr^- cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CGDD is inserted within a marker gene sequence, transformed cells containing sequences encoding CGDD can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CGDD under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CGDD and that express CGDD may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CGDD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CGDD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CGDD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CGDD, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CGDD may be cultured under

conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CGDD may be designed to contain signal sequences which direct secretion of CGDD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CGDD may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CGDD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CGDD activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CGDD encoding sequence and the heterologous protein sequence, so that CGDD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CGDD may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the

T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CGDD of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CGDD. At least one and up to a plurality of test compounds may be screened for specific binding to CGDD. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CGDD, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CGDD binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CGDD, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CGDD or cell membrane fractions which contain CGDD are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CGDD or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CGDD, either in solution or affixed to a solid support, and detecting the binding of CGDD to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CGDD of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CGDD. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CGDD activity, wherein CGDD is combined with at least one test compound, and the activity of CGDD in the presence of a test compound is compared with the activity of CGDD in the absence of the test compound. A change in the activity of CGDD in the presence of the test compound is indicative of a compound that modulates the activity of CGDD. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CGDD under conditions suitable for CGDD activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CGDD may do so indirectly and need not come in direct contact with the test compound.

At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CGDD or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CGDD may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CGDD can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CGDD is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CGDD, e.g., by secreting CGDD in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CGDD and proteins associated with cell growth, differentiation, and death. In addition, examples of tissues expressing CGDD can be found in Table 6. Therefore, CGDD appears to play a role in cell proliferative disorders including cancer, developmental disorders, neurological

disorders, reproductive disorders, and autoimmune/inflammatory disorders. In the treatment of disorders associated with increased CGDD expression or activity, it is desirable to decrease the expression or activity of CGDD. In the treatment of disorders associated with decreased CGDD expression or activity, it is desirable to increase the expression or activity of CGDD.

Therefore, in one embodiment, CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puperty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CGDD in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CGDD may be

administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those listed above.

In a further embodiment, an antagonist of CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD. Examples of such disorders include, but are not limited to, those cell proliferative disorders including cancer, developmental disorders, neurological disorders, reproductive disorders, and autoimmune/inflammatory disorders described above. In one aspect, an antibody which specifically binds CGDD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CGDD.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CGDD may be produced using methods which are generally known in the art. In particular, purified CGDD may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CGDD. Antibodies to CGDD may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with CGDD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol.

Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CGDD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CGDD amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CGDD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CGDD-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CGDD may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CGDD and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CGDD epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CGDD. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CGDD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CGDD epitopes, represents the average affinity, or avidity, of the antibodies for CGDD. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CGDD epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CGDD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CGDD, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CGDD-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CGDD, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CGDD. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CGDD. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CGDD may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in CGDD expression or regulation causes disease, the expression of CGDD from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CGDD are treated by constructing mammalian expression vectors encoding CGDD and introducing these vectors by mechanical means into CGDD-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii)

ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CGDD include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CGDD may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CGDD from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CGDD expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CGDD under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and

A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CGDD to cells which have one or more genetic abnormalities with respect to the expression of CGDD. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CGDD to target cells which have one or more genetic abnormalities with respect to the expression of CGDD. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CGDD to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus

sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CGDD to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CGDD into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CGDD-coding RNAs and the synthesis of high levels of CGDD in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CGDD into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CGDD.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CGDD. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CGDD. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CGDD expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CGDD may be therapeutically useful, and in the treatment of disorders

associated with decreased CGDD expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CGDD may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CGDD is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CGDD are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CGDD. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and

monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CGDD, antibodies to CGDD, and mimetics, agonists, antagonists, or inhibitors of CGDD.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CGDD or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CGDD or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example

CGDD or fragments thereof, antibodies of CGDD, and agonists, antagonists or inhibitors of CGDD, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CGDD may be used for the diagnosis of disorders characterized by expression of CGDD, or in assays to monitor patients being treated with CGDD or agonists, antagonists, or inhibitors of CGDD. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CGDD include methods which utilize the antibody and a label to detect CGDD in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CGDD, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CGDD expression. Normal

or standard values for CGDD expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CGDD under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CGDD expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CGDD may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CGDD may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CGDD, and to monitor regulation of CGDD levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CGDD or closely related molecules may be used to identify nucleic acid sequences which encode CGDD. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CGDD, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CGDD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:13-24 or from genomic sequences including promoters, enhancers, and introns of the CGDD gene.

Means for producing specific hybridization probes for DNAs encoding CGDD include the cloning of polynucleotide sequences encoding CGDD or CGDD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CGDD may be used for the diagnosis of disorders associated with expression of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including

adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of

the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puperty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding CGDD may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CGDD expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CGDD may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CGDD may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CGDD in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CGDD, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CGDD, under conditions suitable for hybridization or

amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CGDD may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CGDD, or a fragment of a polynucleotide complementary to the polynucleotide encoding CGDD, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CGDD may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CGDD are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of CGDD include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used

to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CGDD, fragments of CGDD, or antibodies specific for CGDD may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of

gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CGDD to quantify

the levels of CGDD expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are

well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CGDD may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CGDD on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CGDD, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CGDD and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CGDD, or fragments thereof, and washed. Bound CGDD is then detected by methods well known in the art. Purified CGDD can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CGDD specifically compete with a test compound for binding CGDD. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CGDD.

In additional embodiments, the nucleotide sequences which encode CGDD may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/268,111, U.S. Ser. No. 60/271,175, U.S. Ser. No. 60/274,552, and U.S. Ser. No. 60/274,503, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with

chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in

384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on

Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteins associated with cell growth, differentiation, and death were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteins associated

with cell growth, differentiation, and death, the encoded polypeptides were analyzed by querying against PFAM models for proteins associated with cell growth, differentiation, and death. Potential proteins associated with cell growth, differentiation, and death were also identified by homology to Incyte cDNA sequences that had been annotated as proteins associated with cell growth, differentiation, and death. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended

with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CGDD Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:13-24 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:13-24 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:15 was mapped to chromosome 1 within the interval from

242.50 to 258.70 centiMorgans. SEQ ID NO:20 was mapped to chromosome 7 within the interval from 180.8 centiMorgans to the q-terminus.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding CGDD are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female;

genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CGDD. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CGDD Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in CGDD Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:13-24 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated

algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:13-24 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the CGDD-encoding sequences, or any parts thereof, are used to

detect, decrease, or inhibit expression of naturally occurring CGDD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CGDD. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CGDD-encoding transcript.

XIII. Expression of CGDD

Expression and purification of CGDD is achieved using bacterial or virus-based expression systems. For expression of CGDD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CGDD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CGDD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CGDD by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CGDD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CGDD at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate

resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CGDD obtained by these methods can be used directly in the assays shown in Examples XVII, and XVIII where applicable.

XIV. Functional Assays

CGDD function is assessed by expressing the sequences encoding CGDD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CGDD on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CGDD and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CGDD and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of CGDD Specific Antibodies

CGDD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,

Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CGDD amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CGDD activity by, for example, binding the peptide or CGDD to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring CGDD Using Specific Antibodies

Naturally occurring or recombinant CGDD is substantially purified by immunoaffinity chromatography using antibodies specific for CGDD. An immunoaffinity column is constructed by covalently coupling anti-CGDD antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CGDD are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CGDD (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CGDD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CGDD is collected.

XVII. Identification of Molecules Which Interact with CGDD

CGDD, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CGDD, washed, and any wells with labeled CGDD complex are assayed. Data obtained using different concentrations of CGDD are used to calculate values for the number, affinity, and association of CGDD with the candidate molecules.

Alternatively, molecules interacting with CGDD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially

available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CGDD may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of CGDD Activity

CGDD activity is demonstrated by measuring the induction of terminal differentiation, apoptosis or cell cycle progression when CGDD is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies, Gaithersburg, MD) and PCR 3.1 (Invitrogen, Carlsbad, CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate their physiological state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell cycle progression, cell death or terminal differentiation. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; up or down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

Alternatively, an in vitro assay for CGDD activity measures the transformation of normal human fibroblast cells overexpressing antisense CGDD RNA (Garkavtsev, I. and Riabowol, K. (1997) Mol. Cell Biol. 17:2014-2019). cDNA encoding CGDD is subcloned into the pLNCX retroviral vector to enable expression of antisense CGDD RNA. The resulting construct is transfected into the ecotropic BOSC23 virus-packaging cell line. Virus contained in the BOSC23 culture supernatant is used to infect the amphotropic CAK8 virus-packaging cell line. Virus contained in the

CAK8 culture supernatant is used to infect normal human fibroblast (Hs68) cells. Infected cells are assessed for the following quantifiable properties characteristic of transformed cells: growth in culture to high density associated with loss of contact inhibition, growth in suspension or in soft agar, formation of colonies or foci, lowered serum requirements, and ability to induce tumors when injected into immunodeficient mice. The activity of CGDD is proportional to the extent of transformation of Hs68 cells.

Alternatively, CGDD can be expressed in a mammalian cell line by transforming the cells with a eukaryotic expression vector encoding CGDD. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. To assay the cellular localization of CGDD, cells are fractionated as described by Jiang H. P. et al. (1992; Proc. Natl. Acad. Sci. 89: 7856-7860). Briefly, cells pelleted by low-speed centrifugation are resuspended in buffer (10 mM TRIS-HCl, pH 7.4/ 10 mM NaCl/ 3 mM $MgCl_2$/ 5 mM EDTA with 10 ug/ml aprotinin, 10 ug/ml leupeptin, 10 ug/ml pepstatin A, 0.2 mM phenylmethylsulfonyl fluoride) and homogenized. The homogenate is centrifuged at 600 x g for 5 minutes. The particulate and cytosol fractions are separated by ultracentrifugation of the supernatant at 100,000 x g for 60 minutes. The nuclear fraction is obtained by resuspending the 600 x g pellet in sucrose solution (0.25 M sucrose/ 10 mM TRIS-HCl, pH 7.4/ 2 mM $MgCl_2$) and recentrifuged at 600 x g. Equal amounts of protein from each fraction are applied to an SDS/10% polyacrylamide gel and blotted onto membranes. Western blot analysis is performed using CGDD anti-serum. The localization of CGDD is assessed by the intensity of the corresponding band in the nuclear fraction relative to the intensity in the other fractions. Alternatively, the presence of CGDD in cellular fractions is examined by fluorescence microscopy using a fluorescent antibody specific for CGDD.

Alternatively, CGDD activity may be demonstrated as the ability to interact with its associated Ras superfamily protein, in an in vitro binding assay. The candidate Ras superfamily proteins are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The Ras superfamily proteins are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM MgCl2, 0.2 mM DTT, 100 μM AMP-PNP and 10 μM GDP at 30°C for 20 minutes. CGDD is expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing CGDD-FLAG fusion proteins are precleared with GST beads, then incubated with GST-Ras superfamily fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. CGDD activity is proportional to the amount of CGDD-FLAG fusion protein detected in the complex.

Alternatively, as demonstrated by Li and Cohen (Li, L. and S.N. Cohen (1995) Cell 85:319-

329), the ability of CGDD to suppress tumorigenesis can be measured by designing an antisense sequence to the 5' end of the gene and transfecting NIH 3T3 cells with a vector transcribing this sequence. The suppression of the endogenous gene will allow transformed fibroblasts to produce clumps of cells capable of forming metastatic tumors when introduced into nude mice.

Alternatively, an assay for CGDD activity measures the effect of injected CGDD on the degradation of maternal transcripts. Procedures for oocyte collection from Swiss albino mice, injection, and culture are as described in Stutz (supra). A decrease in the degradation of maternal RNAs as compared to control oocytes is indicative of CGDD activity. In the alternative, CGDD activity is measured as the ability of purified CGDD to bind to RNAse as measured by the assays described in Example XVII.

Alternatively, an assay for CGDD activity measures syncytium formation in COS cells transfected with an CGDD expression plasmid, using the two-component fusion assay described in Mi (supra). This assay takes advantage of the fact that human interleukin 12 (IL-12) is a heterodimer comprising subunits with molecular weights of 35 kD (p35) and 40 kD (p40). COS cells transfected with expression plasmids carrying the gene for p35 are mixed with COS cells cotransfected with expression plasmids carrying the genes for p40 and CGDD. The level of IL-12 activity in the resulting conditioned medium corresponds to the activity of CGDD in this assay. Syncytium formation may also be measured by light microscopy (Mi et al. supra).

An alternative assay for CGDD activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CGDD is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine or a radioactive DNA precursor such as [α^{32}P]ATP. Where applicable, varying amounts of CGDD ligand are added to the transfected cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA and CGDD activity.

Alternatively, CGDD activity is measured by the cyclin-ubiquitin ligation assay (Townsley, F.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2362-2367). The reaction contains in a volume of 10 μl, 40 mM Tris.HCl (pH 7.6), 5 mM Mg Cl_2, 0.5 mM ATP, 10 mM phosphocreatine, 50 μg of creatine phosphokinase/ml, 1 mg reduced carboxymethylated bovine serum albumin/ml, 50 μM ubiquitin, 1 μM ubiquitin aldehyde, 1-2 pmol ^{125}I-labeled cyclin B, 1 pmol E1, 1 μM okadaic acid, 10 μg of protein of M-phase fraction 1A (containing active E3-C and essentially free of E2-C), and varying amounts of CGDD. The reaction is incubated at 18 °C for 60 minutes. Samples are then separated by electrophoresis on an SDS polyacrylamide gel. The amount of ^{125}I- cyclin-ubiquitin formed is quantified by PHOSPHORIMAGER analysis. The amount of cyclin-ubiquitin formation is

proportional to the activity of CGDD in the reaction.

Alternatively, an assay for CGDD activity uses radiolabeled nucleotides, such as [α^{32}P]ATP, to measure either the incorporation of radiolabel into DNA during DNA synthesis, or fragmentation of DNA that accompanies apoptosis. Mammalian cells are transfected with plasmid containing cDNA encoding CGDD by methods well known in the art. Cells are then incubated with radiolabeled nucleotide for various lengths of time. Chromosomal DNA is collected, and radioactivity is detected using a scintillation counter. Incorporation of radiolabel into chromosomal DNA is proportional to the degree of stimulation of the cell cycle. To determine if CGDD promotes apoptosis, chromosomal DNA is collected as above, and analyzed using polyacrylamide gel electrophoresis, by methods well known in the art. Fragmentation of DNA is quantified by comparison to untransfected control cells, and is proportional to the apoptotic activity of CGDD.

Alternatively, cyclophilin activity of CGDD is measured using a chymotrypsin-coupled assay to measure the rate of cis to trans interconversion (Fischer, G., Bang, H., and Mech, C. (1984) Biomed. Biochim. Acta 43: 1101-1111). The chymotrypsin is used to estimate the trans-substrate cleavage activity at Xaa-Pro peptide bonds, wherein the rate constant for the cis to trans isomerization can be obtained by measuring the rate constant of the substrate hydrolysis at the slow phase. Samples are incubated in the presence or absence of the immunosuppressant drugs CsA or FK506, reactions initiated by addition of chymotrypsin, and the fluorescent reaction measured. The enzymatic rate constant is calculated from the equation $k_{app} = k_{H2O} + k_{enz}$, wherein first order kinetics are displayed, and where one unit of PPIase activity is defined as k_{enz} (s^{-1}).

Alternatively, cyclophilin activity of CGDD is monitored by a quantitative immunoassay that measures its affinity for stereospecific binding to the immunosuppressant drug cyclosporin (Quesniaux, V.F. et al. (1987) Eur. J. Immunol. 17: 1359-1365). In this assay, the cyclophilin-cyclosporin complex is coated on a solid phase, with binding detected using anti-cyclophilin rabbit antiserum enhanced by an antiglobulin-enzyme conjugate.

Alternatively, activity of CGDD is monitored by a binding assay developed to measure the non-covalent binding between FKBPs and immunosuppressant drugs in the gas phase using electrospray ionization mass spectrometry (Trepanier, D.J., et al. (1999) Ther. Drug Monit. 21: 274-280). In electrospray ionization, ions are generated by creating a fine spray of highly charged droplets in the presence of a strong electric field; as the droplet decreases in size, the charge density on the surface increases. Ions are electrostatically directed into a mass analyzer, where ions of opposite charge are generated in spatially separate sources and then swept into capillary inlets where the flows are merged and where reactions occur. By comparing the charge states of bound versus unbound CGDD/immunosuppressive drug complexes, relative binding affinities can be established and correlated with <u>in vitro</u> binding and immunosuppressive activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1567742	1	1567742CD1	13	1567742CB1
7485501	2	7485501CD1	14	7485501CB1
3089944	3	3089944CD1	15	3089944CB1
5284076	4	5284076CD1	16	5284076CB1
2899903	5	2899903CD1	17	2899903CB1
7491355	6	7491355CD1	18	7491355CB1
3333288	7	3333288CD1	19	3333288CB1
7488313	8	7488313CD1	20	7488313CB1
6013113	9	6013113CD1	21	6013113CB1
7488573	10	7488573CD1	22	7488573CB1
7506027	11	7506027CD1	23	7506027CB1
7503618	12	7503618CD1	24	7503618CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	1567742CD1	g7144644	0.0	[Homo sapiens] tumor antigen SLP-8p
2	7485501CD1	g16876842	1.0E-47	[Homo sapiens] tumor suppressor deleted in oral cancer-related 1
2	7485501CD1	g3661529	6.3E-45	[Homo sapiens] growth suppressor related Zhang, X., et al. (1999) Biochem. Biophys. Res. Commun. 255:59-63
3	3089944CD1	g5114353	2.2E-102	[Rattus norvegicus] RING finger protein terf Ogawa, S. et al. (1998) Biochem. Biophys. Res. Commun. 251:515-519
4	5284076CD1	g8917577	1.0E-19	[Mus musculus] EPCS26 Hemberger, M.C., et al. Dev. Biol. (2000) 222:158-69
5	2899903CD1	g4105589	0.0	[Homo sapiens] nGAP Noto, S. et al. (1998) FEBS Lett. 441:127-131
6	7491355CD1	g2204355	1.6E-144	[Mus musculus] radical fringe (boundary determination/ Notch pathway) precursor Johnston, S.H. et al. (1997) Development 124:2245-2254
7	3333288CD1	g440127	1.4E-260	[Rattus norvegicus] cerebroglycan (neuronal differentiation associated) Stipp, C.S. et al. (1994) J. Cell Biol. 124:149-160
8	7488313CD1	g9651220	5.9E-248	[Mus musculus] LMBR1 (polydactyly associated) long form Clark, R.M. et al. (2000) Genomics 67:19-27
9	6013113CD1	g339973	3.9E-244	[Homo sapiens] TRPM-2 gene product (Clusterin) Wong, P. et al. (1993) J. Biol. Chem. 268:5021-5031
10	7488573CD1	g3170615	0.0	[Mus musculus] DOC4 Wang, X.Z. et al. (1998) EMBO J. 17:3619-3630
11	7506027CD1	g4105589	0.0	[Homo sapiens] nGAP Noto, S. et al. (1998) supra
12	7503618CD1	g339973	0.0	[Homo sapiens] TRPM-2 gene product (Clusterin) Wong, P. et al. (1993) supra

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1567742CD1	977	S14 S35 S87 S125 S155 S201 S225 S237 S249 S308 S335 S362 S422 S478 S536 S551 S553 S563 S605 S707 S810 S858 S867 S874 S920 T157 T321 T408 T588 T658 T803 Y779	N918	Transmembrane domain: E725-I752	TMAP
					PROTEIN C2F3.10 CHROMOSOME I T21C9.2 PD025207: L732-L932	BLAST_PRODOM
2	7485501CD1	109	S2 S97 T73		SUPPRESSOR PUTATIVE ORAL CANCER DELETED CANCER1 ANTI-ONCOGENE DOC1 GROWTH RELATED PD020621: S11-H108	BLAST_PRODOM
3	3089944CD1	468	S173 S339 S390 S419 T281 T437	N388	SPRY domain: S339-K458	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C16-C56	HMMER_PFAM
					B-box zinc finger: V87-L128	HMMER_PFAM
					Transmembrane domains: E370-Y387 H414-F442	TMAP
					Zinc finger, C3HC4 type (RING finger), signature: L10-R63	PROFILESCAN
					Domain in SPla and the Ryanodine receptor PF00622B: E323-W344, V404-F417	BLIMPS_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3 (cont.)					ZINC FINGER RFP FINGER RET METAL BINDING NUCLEAR DNA BINDING SIMILAR THE PD032801: Q129-R270, E233-F337	BLAST_PRODOM
					BUTYROPHILIN ZINC FINGER NUCLEAR FINGER DNA BINDING RET RNA BINDING PRECURSOR BT PD002445: E233-F337	BLAST_PRODOM
					FINGER MIDLINE ZINC FINGER RING STONUSTOXIN PUTATIVE TRANSCRIPTION FACTOR XPRF PD002421: L291-T453	BLAST_PRODOM
					DOWN REGULATORY PROTEIN OF INTERLEUKIN 2 RECEPTOR TRANSCRIPTION REGULATION DNA BINDING TRANSACTING FACTOR ZINC FINGER PD084482: E133-G287	BLAST_PRODOM
					RFP TRANSFORMING PROTEIN DM02346|P19474|59-337: R63-F337 DM02346|P14373|61-366: P61-F285, D288-F337 DM01944|P18892|355-477: S339-C455 DM02346|A57041|64-348: N64-Q310	BLAST_DOMO
					Cell attachment sequence: R286-D288, R311-D313	MOTIFS
					Zinc finger, C3HC4 type (RING finger), signature: C31-I40	MOTIFS
					Leucine zipper pattern: L211-L232	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	5284076CD1	158	S115 T15 T97 T114 T126 T143 T144		signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-N19	HMMER
					Chromo domain signature and profile: R99-G149	PROFILESCAN
5	2899903CD1	1161	S5 S15 S22 S29 S31 S52 S58 S73 S114 S122 S155 S158 S180 S340 S371 S432 S471 S477 S485 S541 S568 S601 S700 S719 S814 S821 S833 S867 S936 S945 S946 S1003 S1008 S1024 S1080 T218 T251 T343 T358 T519 T771 T849 T869 T1025 T1042 Y262 Y283	N71 N1157	PH domain: E126-H174	HMMER_PFAM
					GTPase-activator protein for Ras-like G: F364-F536	HMMER_PFAM
					Transmembrane domain: S485-L507 N-terminus is cytosolic	TMAP
					Ras GTPase-activating proteins signature and profile ras_gtpase_activ.prf: L398-L525	PROFILESCAN
					Ras GTPase-activating protein BL00509B:L525-N535	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5 (cont.)					GAP24 PD142012: P35-F364	BLAST_PRODOM
					PROTEIN GTPASE ACTIVATION GTPASE ACTIVATING RAS NEUROFIBROMIN P21 ACTIVATOR INHIBITORY REGULATOR PD002301: L436-N535	BLAST_PRODOM
					RAS-SPECIFIC GAP CATALYTIC DOMAIN DM08490|B40121|268-786: L119-E554	BLAST_DOMO
					RAS-SPECIFIC GAP CATALYTIC DOMAIN DM08490|P09851|442-960: L119-E554	BLAST_DOMO
6	7491355CD1	331	S54 S137 S227 S260 T66 T67 T94 T183 T194 T329	N113	Signal Peptide: M1-A21, M1-P33	HMMER
					Fringe-like: P53-R305	HMMER_PFAM
					Predicted transmembrane segments: S2-P30, V195-W219; N-terminus non-cytosolic	TMAP
					FRINGE PRECURSOR SIGNAL DEVELOPMENTAL TRANSFERASE PD005426: P39-K324	BLAST_PRODOM
					Leucine zipper pattern: L8-L29	MOTIFS
7	3333288CD1	579	S31 S71 S197 S209 S285 S335 S446 S461 S543 T84 T162 T233 T374 T467 T478 Y330		Signal Peptide: M1-G19, M1-G21, M1-P20, M1-E25, M1-V28	HMMER
					Glypican: A3-L566	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7 (cont.)					Glypicans proteins BL01207: C62-L77, C191-N236, C250-S285, C429-P463, G487-G503	BLIMPS_BLOCKS
					PRECURSOR PROTEOGLYCAN HEPARAN SULFATE GLYCOPROTEIN SIGNAL GPI ANCHOR PROTEIN GLYPICAN1 EXTRACELLULAR PD007065: N142-P527, L8-Y152	BLAST_PRODOM
					GLYPICAN; DM03626 P51653|1-578: M1-R579 P51655|1-556: L7-G505, D553-Q548, D489-V517 P35052|1-557: G23-P527 P50593|1-549: L14-I558	BLAST_DOMO
					Glypicans signature: C260-C283	MOTIFS
8	7488313CD1	490	S8 S46 S181 S256 S286 S368 T228 T241 T274 T277 T376 T411 T451 T472 T478	N101 N337 N471	Predicted transmembrane segments: R58-E86, G107-L131, G142-W166, Y189-F217, N291-L319, L340-S368, T385-F413, G420-R448	TMAP
					HYPOTHETICAL 56.4 KD PROTEIN PD142903: L61-G221, I23-D47	BLAST_PRODOM
					PROTEIN R05D3.2 CHROMOSOME III PD025307: L61-V225, L223-E457	BLAST_PRODOM
9	6013113CD1	544	S27 S72 S161 S185 S233 S357 S394 S424 S455 S540 T25 T58 T63 T338 T376 T433 T459 Y206	N86 N103 N145 N291 N354 N374	Signal peptide: M1-Q24, M2-Q19, M2-D23, M2-Q24	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 (cont.)					Clusterin: M2-S449	HMMER-PFAM
					Clusterin proteins BL00492: M2-G18, V26-N48, G52-L85, N86-M122, V128-M176, F218-D259, C285-A334, D413-S449	BLIMPS-BLOCKS
					Clusterin signatures: T93-E141, I275-R325	ProfileScan
					Clusterin, glycoprotein, signal protein, plasma complement, cytolysis inhibitor PD006991:M2-D279, Q168-D448	BLAST-PRODOM
					Clusterin: DM07724|P17697|1-438: M2-D448 DM07724|P14018|1-450: L5-H446	BLAST-DOMO
					Clusterin signature 1: C113-C121	MOTIFS
					Clusterin signature 2: C295-C305	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7488573CD1	2758	S135 S138 S145 S272 S332 S425 S439 S520 S652 S664 S889 S950 S1030 S1061 S25 S1140 S1210 S40 S1288 S1378 T13 S1386 S1500 T23 S1514 S1569 T44 S1624 S1704 T79 S1707 S1720 S1721 S1756 S1866 S1891 S1928 S2077 S2103 S2195 S2220 S2277 S2290 S2319 S2340 S2341 S2614 T153 S113 T155 T157 T204 T258 T400 T485 T509 T554 T683 T687 T691 T716 T739 T765 T828 T1152 T1488 T1581 T1682 T1836 T1838 T1849 T1903 T1916 T1957 T1995 T2012 T2016 T2022 T2043 T2178 T2418 T2545 T2649 T2653 T2670 T2671	N77 N151 N463 N936 N1255 N1598 N1694 N1730 N1788 N1873 N1974 N2177 N2317 N2635	EGF-like domain: C756-C782, C626-C653, C658-C685, C796-C826, C725-C751, C692-C720, C562-C588, C593-C619	HMMER-PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10 (cont.)			T2710 T2720 Y22 Y2014 Y2145 Y2242 Y2260		NHL repeat: L1395-V1430, L1525-F1551	HMMER-PFAM
					Transmembrane domains: W337-N365, T1337-R1360, E2344-K2367 N-terminus is cytosolic	TMAP
					Type III EGF-like signature PR00011: G570-C588, G764-C782	BLIMPS-PRINTS
					ODD, OZ, tenascin-like, DOC4, glycoprotein PD011966:P931-T1645, N1189-G1811, R1805-S2077, Y1968-A2304, Y2185-I2234, D2256-F2335, W2154-G2264, T1663-T1709, N151-T186, I1607-I1751, Y2181-L2214, Y1932-P1948, S348-V388, E1895-Y1915, L1987-Q2036, G1603-R1621	BLAST-PRODOM
					ODD, OZ, tenascin-like, DOC4, glycoprotein PD018620: P2309-R2758	BLAST-PRODOM
					DOC4, glycoprotein PD185998: N2076-N2308, G2580-E2594	BLAST-PRODOM
					Gammaheregulin DOC4 PD151529: P165-K410, D2-P180	BLAST-PRODOM
					EGF DM00003| P24821|206-292: C711-C806, C579-C669, C562-C637, I643-C735 A45445|178-268: H699-C782, Y567-C658, C674-C756, C642-D728 S47008|430-483: C692-E743	BLAST-DOMO
					Tenascin DM05547:S47008|565-645: C834-S902	BLAST-DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10 (cont.)					EGF-like domain signature 1: C577-C588, C608-C619, C642-C653, C674-C685, C709-C720, C740-C751, C771-C782, C815-C826	MOTIFS
					EGF-like domain signature 2: C577-C588, C608-C619, C642-C653, C674-C685, C740-C751, C771-C782, C815-C826	MOTIFS
11	7506027CD1	1139	S5 S15 S22 S29 S31 S52 S58 S73 S114 S122 S155 S158 S180 S340 S371 S432 S471 S477 S485 S541 S568 S601 S678 S697 S792 S799 S811 S845 S914 S923 S924 S981 S986 S1002 S1058 T218 T251 T343 T358 T519 T749 T827 T847 T1003 T1020 Y262 Y283	N71 N1135	PH domain: E126-H174	HMMER_PFAM
					GTPase-activator protein for Ras-like GTPase: F364-F536	HMMER_PFAM
					Ras GTPase-activating proteins BL00509: L525-N535	BLIMPS_BLOCKS
					Ras GTPase-activating proteins signature and profile: L398-L525	PROFILESCAN
					GAP24 PD142012: P35-F364	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11 (cont.)					PROTEIN GTPASE ACTIVATION GTPASE-ACTIVATING RAS NEUROFIBROMIN P21 ACTIVATOR INHIBITORY REGULATOR PD002301: L96-I136, L355-Q440, L436-N535	BLAST_PRODOM
					RAS-SPECIFIC GAP CATALYTIC DOMAIN DM08490|B40121|268-786: L119-E554 DM08490|P09851|442-960: L119-E554	BLAST_DOMO
					Leucine zipper pattern: L985-L1006	MOTIFS
12	7503618CD1	503	S27 S72 S161 S185 S192 S316 S353 S383 S414 S499 T25 T58 T63 T297 T335 T392 T418	N86 N103 N145 N250 N313 N333	signal_cleavage: M1-G18	SPSCAN
					Clusterin: M2-F191, S192-S408	HMMER_PFAM
					Signal Peptide: M2-S17, M2-V20, M2-G22, M1-G22	HMMER
					Clusterin proteins BL00492: M2-G18, V26-N48, G52-L85, N86-M122, V128-M176, Q177-D218, C244-A293, D372-S408	BLIMPS_BLOCKS
					Clusterin signatures 1: T93-E141	PROFILESCAN
					Clusterin signatures 2: I234-R284	PROFILESCAN
					PRECURSOR GLYCOPROTEIN CLUSTERIN SIGNAL PROTEIN PLASMA COMPLEMENT CYTOLYSIS INHIBITOR CLI PD006991: M2-F206, F191-D407	BLAST_PRODOM

Table 3

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>12 (cont.)</td><td></td><td></td><td></td><td></td><td>CLUSTERIN
DM07724|P17697|1-438: M2-F191, F191-D407
DM07724|P14018|1-450: L5-R235, P196-H405</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Clusterin signature 1: C113-C121</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Clusterin signature 2: C254-C264</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
13/1567742CB1/3971	1-786, 311-647, 373-826, 391-733, 486-3485, 677-1124, 764-1045, 764-1360, 883-994, 883-1211, 883-1407, 909-1133, 1113-1400, 1377-1553, 1515-1620, 1515-1641, 1519-1640, 1586-1754, 1640-2207, 1846-2174, 1882-1903, 1910-2638, 1928-2164, 2023-2110, 2078-2365, 2232-2430, 2242-2668, 2244-2448, 2244-2668, 2292-2517, 2320-2609, 2324-2854, 2339-2592, 2378-2944, 2412-3063, 2451-2958, 2455-2734, 2462-2918, 2464-2679, 2468-2772, 2499-2784, 2532-3121, 2532-3148, 2580-2918, 2607-2894, 2652-3148, 2657-3135, 2657-3145, 2657-3148, 2658-3145, 2658-3148, 2660-2857, 2660-3148, 2663-3145, 2663-3148, 2667-2931, 2669-2946, 2669-2998, 2669-3039, 2669-3148, 2676-2838, 2704-2843, 2704-2932, 2704-3133, 2704-3147, 2704-3199, 2704-3245, 2704-3295, 2704-3312, 2704-3320, 2704-3358, 2704-3364, 2704-3367, 2704-3383, 2704-3414, 2704-3562, 2732-2897, 2732-2911, 2732-3149, 2732-3168, 2743-2932, 2771-3000, 2778-3209, 2781-3049, 2836-3094, 2837-3056, 2871-3450, 2885-3313, 2895-3396, 2905-3403, 2906-3150, 2937-3373, 2980-3540, 2983-3208, 2986-3524, 3023-3218, 3026-3549, 3033-3277, 3039-3524, 3040-3342, 3049-3649, 3090-3365, 3098-3390, 3104-3249, 3104-3351, 3105-3786, 3151-
	3626, 3173-3711, 3173-3777, 3188-3458, 3190-3807, 3206-3807, 3211-3827, 3214-3798, 3227-3551, 3233-3520, 3233-3523, 3233-3695, 3238-3618, 3248-3802, 3258-3464, 3263-3518, 3271-3452, 3289-3807, 3297-3503, 3304-3950, 3314-3570, 3317-3946, 3326-3957, 3332-3831, 3332-3911, 3332-3967, 3337-3559, 3342-3610, 3342-3659, 3342-3762, 3342-3812, 3355-3795, 3363-3812, 3363-3889, 3367-3660, 3371-3673, 3372-3792, 3385-3656, 3412-3643, 3412-3966, 3412-3967, 3413-3688, 3413-3957, 3430-3681, 3440-3660, 3442-3809, 3451-3712, 3465-3695, 3492-3957, 3506-3971, 3515-3779, 3528-3785, 3528-3796, 3530-3966, 3532-3797, 3536-3921, 3550-3837, 3550-3949, 3552-3967, 3576-3971, 3650-3887, 3650-3910, 3650-3912, 3650-3934, 3656-3809, 3668-3958, 3678-3877, 3695-3940, 3728-3966, 3809-3966
14/7485501CB1/410	1-383, 15-386, 44-234, 44-410, 65-338, 148-329, 200-395, 200-399, 200-410

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
15/3089944CB1/2597	1-424, 126-404, 126-541, 251-943, 500-1021, 533-980, 533-1021, 534-744, 534-812, 534-1117, 534-1123, 534-1182, 535-758, 550-831, 558-722, 561-750, 566-1195, 576-1046, 576-1104, 578-1046, 584-850, 584-885, 588-1193, 591-924, 638-1192, 719-973, 732-1012, 800-1452, 1007-1288, 1011-1197, 1185-1763, 1203-1706, 1253-1530, 1299-1531, 1299-1822, 1330-1618, 1334-1652, 1335-1582, 1336-1678, 1348-1604, 1383-1756, 1556-1800, 1572-1837, 1596-1872, 1607-1885, 1607-1886, 1614-1846, 1621-1917, 1636-1909, 1636-2253, 1643-1857, 1645-1964, 1658-1914, 1658-1924, 1658-1942, 1675-1953, 1739-1959, 1789-2026, 1824-2072, 1872-2088, 1873-2175, 1880-2116, 1880-2390, 1915-2197, 1926-2203, 1956-2169, 1956-2281, 1992-2275, 2008-2586, 2049-2333, 2075-2597, 2089-2586, 2123-2360, 2129-2334, 2169-2413, 2169-2543, 2171-2579, 2172-2417, 2172-2427, 2172-2597, 2192-2544, 2195-2560, 2213-2578, 2240-2597, 2258-2534, 2346-2597, 2452-2597, 2459-2597
16/5284076CB1/1480	1-174, 1-521, 1-605, 7-537, 29-274, 108-375, 115-375, 185-603, 206-657, 295-603, 479-1272, 596-1223, 686-1479, 781-1444, 800-1409, 835-1474, 847-1452, 865-1342, 961-1480, 1076-1446
17/2899903CB1/6877	1-605, 64-659, 242-754, 272-522, 392-772, 392-779, 394-495, 394-730, 394-744, 495-1175, 516-1226, 554-1132, 597-1272, 599-1196, 625-912, 628-3837, 629-893, 629-1015, 837-1581, 856-1460, 960-1766, 1064-1795, 1071-1795, 1073-1795, 1103-1795, 1104-1740, 1104-1780, 1118-1445, 1118-1583, 1121-1706, 1121-1779, 1123-1795, 1134-1795, 1138-1795, 1141-1795, 1149-1795, 1171-1795, 1173-1795, 1176-1795, 1179-1795, 1184-1795, 1206-1686, 1207-1795, 1215-1795, 1225-1795, 1228-1795, 1232-1795, 1235-1533, 1241-1795, 1250-1795, 1257-1795, 1269-1795, 1308-1795, 1312-1795, 1342-1795, 1366-1795, 1383-1795, 1481-2145, 1537-1859, 1728-1785, 1781-2292, 1783-2589, 1787-2322, 1792-2439, 1793-2345, 1795-2345, 1796-2345, 1860-2345, 1912-2330, 2059-2710, 2236-2749, 2951-3555, 2994-3573, 3045-3590, 3064-3617, 3065-3288, 3066-3244, 3068-3315, 3202-3806, 3249-3748, 3261-3846, 3287-3539, 3388-3758, 3420-3891, 3425-3697, 3425-3979, 3493-3990, 3515-4302, 3545-3963, 3549-3991, 3551-3963, 3560-3963, 3566-3740, 3568-3989, 3572-3963, 3600-4313, 3627-3991, 3670-3991, 3675-

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
17 (cont.)	3933, 3830-3987, 3983-4151, 3983-4451, 3983-4552, 3983-4590, 3983-4604, 3984-4366, 3984-4424, 3984-4464, 3985-4451, 4050-4677, 4060-4516, 4060-4586, 4182-4455, 4182-4717, 4184-4769, 4196-4782, 4221-4523, 4226-4473, 4245-4845, 4289-4794, 4355-4604, 4358-5080, 4586-4825, 4586-5053, 4623-5176, 4623-5272, 4683-4971, 4753-4998, 4753-5015, 4772-5058, 4808-5387, 4873-5111, 4873-5216, 4890-5163, 4959-5153, 4959-5442, 4960-5274, 4963-5585, 4968-5263, 4983-5599, 5201-5465, 5217-5494, 5218-5502, 5312-5616, 5343-5599, 5382-5667, 5429-5581, 5431-5704, 5446-5727, 5477-5694, 5484-5668, 5484-5736, 5484-5751, 5487-5682, 5617-5933, 5623-5863, 5623-5914, 5628-5879, 5644-5903, 5668-5901, 5681-5950, 5746-6026, 5761-6043, 5844-5987, 5844-6374, 5877-6123, 5879-6115, 5892-6190, 5917-6199, 5923-6127, 5923-6142, 5943-6116, 5967-6218, 5971-6159, 5977-6205, 5997-6278, 6056-6284, 6073-6282, 6073-6303, 6074-6319, 6074-6606, 6114-6352, 6115-6354, 6151-6387, 6168-6859, 6174-6438, 6176-6396, 6179-6407, 6179-6453, 6183-6405, 6205-6403, 6210-6853, 6243-6853, 6284-6528, 6284-6809, 6328-6575, 6349-6595, 6389-6847, 6401-6656, 6431-6686, 6442-6664, 6444-6684, 6444-6857,
	6444-6877, 6445-6696, 6600-6847, 6693-6828
18/7491355CB1/1290	1-104, 1-114, 16-224, 17-225, 165-448, 265-566, 265-746, 291-536, 300-545, 300-708, 303-614, 313-933, 316-662, 316-859, 321-894, 326-949, 334-606, 335-626, 351-638, 363-936, 374-541, 411-985, 416-999, 421-1154, 433-680, 461-658, 469-721, 478-1092, 510-799, 558-790, 563-652, 563-1148, 572-841, 574-662, 613-1093, 632-930, 651-834, 653-912, 659-761, 659-772, 659-881, 675-938, 688-935, 693-960, 696-971, 730-1003, 732-1027, 738-1045, 757-1100, 786-1094, 789-1073, 806-1042, 847-1123, 852-1106, 852-1153, 887-1216, 943-1290, 944-1217, 944-1228, 954-1244, 982-1243, 982-1289, 1006-1260
19/3333288CB1/2133	1-474, 4-471, 4-474, 8-470, 8-474, 10-550, 11-474, 11-550, 14-550, 22-148, 23-474, 26-472, 26-474, 27-550, 30-474, 34-465, 344-994, 347-497, 373-620, 422-1119, 589-1217, 920-1247, 1065-1625, 1413-1918, 1413-2131, 1413-2133, 1417-2133, 1476-2133, 1539-2133

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
20/7488313CB1/5162	1-311, 2-577, 8-446, 362-388, 362-701, 362-788, 362-931, 430-917, 444-1083, 500-1014, 503-1014, 568-587, 579-1164, 654-971, 696-1315, 704-943, 706-983, 765-1020, 765-1228, 781-1401, 832-1431, 849-1390, 861-1437, 861-1442, 880-1522, 884-1579, 898-1291, 922-1569, 941-1220, 991-1558, 1003-1448, 1098-1712, 1108-1830, 1150-1350, 1152-1368, 1152-1626, 1152-1787, 1159-1524, 1174-1805, 1175-1867, 1304-1570, 1374-1580, 1411-1719, 1492-1650, 1492-1727, 1493-2028, 1521-1761, 1550-2095, 1615-2162, 1750-2158, 1750-2162, 1750-2266, 1816-2001, 1818-2162, 1818-2207, 1818-2255, 1818-2263, 1818-2287, 1818-2290, 1818-2301, 1820-2198, 1829-2255, 1860-2250, 1894-2549, 1929-2088, 1929-2212, 1929-2220, 1929-2229, 1944-2201, 1973-2555, 2043-2516, 2089-2335, 2095-2458, 2096-2380, 2131-2697, 2134-2370, 2143-2541, 2148-2460, 2156-2649, 2163-2555, 2168-2462, 2168-2493, 2168-2682, 2203-2782, 2216-2459, 2234-2500, 2245-2499, 2245-2702, 2285-2668, 2332-2793, 2388-2618, 2388-2813, 2388-2897, 2394-2639, 2431-3045, 2463-2747, 2480-2793, 2483-2792, 2492-2750, 2493-2649, 2528-2792, 2533-3065, 2563-2804, 2578-2782, 2598-3190, 2600-2792, 2633-2943, 2651-2908, 2677-3139, 2678-
	3297, 2741-3153, 2746-3104, 2791-3302, 2860-3166, 2970-3302, 2996-3249, 3007-3273, 3053-3640, 3066-3271, 3068-3307, 3088-3325, 3088-3372, 3096-3336, 3096-3457, 3096-3661, 3105-3512, 3105-3542, 3105-3649, 3148-3588, 3155-3302, 3181-3435, 3188-3718, 3271-3770, 3322-3589, 3322-3802, 3383-3943, 3398-3984, 3400-3754, 3460-3951, 3544-4197, 3555-4192, 3565-4201, 3577-3915, 3586-4018, 3598-3861, 3660-4213, 3730-3972, 3738-4171, 3775-4210, 3801-4205, 3802-4210, 3804-4209, 3810-4203, 3813-4204, 3872-4126, 3872-4349, 3885-4209, 3908-4172, 3921-4210, 3969-4147, 4139-4418, 4219-4490, 4219-4491, 4219-4531, 4247-4513, 4247-4814, 4278-4839, 4418-4646, 4418-4814, 4418-4848, 4418-4852, 4524-5162

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
21/6013113CB1/1712	1-1712, 49-525, 49-527, 49-534, 49-538, 49-539, 49-541, 49-542, 49-543, 49-544, 49-545, 49-550, 49-552, 49-555, 49-558, 49-560, 49-561, 49-563, 49-567, 49-569, 49-571, 49-574, 49-575, 49-576, 49-578, 49-579, 49-581, 49-582, 49-585, 49-586, 49-588, 49-591, 49-593, 49-597, 49-599, 49-600, 49-602, 49-604, 49-606, 49-611, 49-613, 49-616, 49-619, 49-622, 49-624, 49-626, 49-628, 49-629, 49-631, 49-632, 49-633, 49-635, 49-638, 49-640, 49-641, 49-643, 49-645, 49-649, 49-650, 49-653, 49-654, 49-655, 49-659, 49-660, 49-661, 49-666, 49-668, 49-671, 49-672, 49-673, 49-674, 49-676, 49-677, 49-678, 49-682, 49-683, 49-700, 49-721, 49-732, 49-739, 49-747, 49-788, 49-803, 49-804, 49-889, 49-896, 50-660, 50-674, 50-677, 51-533, 51-701, 53-578, 53-621, 53-704, 61-674, 64-661, 66-606, 72-675, 73-698, 75-660, 75-700, 76-862, 80-659, 85-654, 89-604, 93-691, 94-678, 97-595, 97-619, 100-618, 103-528, 103-723, 105-574, 105-807, 108-619, 116-691, 117-566, 118-686, 121-666, 123-705, 125-802, 129-811, 130-549, 130-587, 135-798, 135-907, 140-738, 140-761, 140-841, 148-787, 151-717, 152-246, 153-802, 159-614, 160-836, 160-879, 161-748, 161-771, 164-623, 165-551, 165-827, 168-722, 172-745, 173-779, 177-698, 177-809, 179-639, 179-874, 187-664, 192-847, 194-844, 195-827, 197-659, 197-738, 200-738, 200-773, 200-923, 202-898, 207-617, 208-750, 208-755, 208-778, 209-538, 209-773, 218-732, 218-884, 227-776, 228-830, 230-832, 231-834, 233-838, 233-848, 236-909, 238-910, 240-735, 241-720, 241-817, 252-710, 258-657, 258-781, 269-878, 271-879, 274-683, 275-746, 275-791, 275-842, 276-784, 276-857, 277-804, 277-863, 283-711, 285-664, 285-674, 300-944, 307-722, 307-737, 307-760, 307-785, 308-870, 308-910, 308-995, 316-1020, 317-731, 317-828, 317-873, 320-898, 322-884, 335-672, 335-825, 340-525
22/7488573CB1/8645	1-523, 339-859, 348-587, 432-523, 608-859, 609-1349, 958-1057, 1211-1392, 1211-1393, 1223-1392, 1321-1994, 1380-1982, 1584-1905, 1584-2019, 1584-2038, 1584-2284, 1584-2341, 2026-2272, 2026-2322, 2033-2840, 2046-2842, 2081-2285, 2161-2983, 2164-2646, 2209-2469, 2209-2807, 2211-2469, 2296-3106, 2381-3116, 2411-2468, 2416-2873, 2450-2916, 2452-3239, 2460-2662, 2496-3032, 2577-3163, 2587-3399, 2639-3438, 2734-3467, 2783-3645, 2792-3580, 2808-3321, 2819-3342, 2933-3622, 3204-3821, 3352-3904, 3355-3904, 3446-3904, 3511-4292, 3701-3729, 4046-4510, 4070-4742, 4137-8390, 4218-4739, 4293-5342, 7555-8347, 7621-8107, 7622-8195, 7631-8109, 7632-8109, 7635-7964, 7635-8046, 7676-8109, 7778-7963, 7778-8303, 7842-8482, 7875-8470, 8012-8486, 8062-8306, 8215-8645

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
23/7506027CB1/6812	1-605, 1-6812, 64-659, 242-754, 272-522, 272-589, 495-1175, 554-1132, 563-637, 599-1196, 625-912, 629-893, 629-1015, 643-1314, 645-1107, 652-1068, 702-952, 719-910, 719-1058, 736-1019, 837-1574, 856-1202, 856-1460, 898-1110, 960-1766, 1064-1785, 1071-1785, 1073-1785, 1103-1785, 1104-1740, 1104-1780, 1118-1445, 1118-1583, 1121-1706, 1121-1779, 1123-1785, 1130-1785, 1134-1785, 1138-1785, 1141-1785, 1149-1788, 1171-1785, 1173-1785, 1176-1785, 1179-1785, 1184-1785, 1207-1785, 1215-1785, 1225-1785, 1228-1785, 1232-1785, 1236-1533, 1241-1785, 1250-1785, 1252-1785, 1256-1785, 1257-1785, 1258-1788, 1269-1785, 1279-1785, 1295-1745, 1300-1745, 1308-1785, 1312-1785, 1321-1745, 1342-1785, 1383-1785, 1384-1786, 1393-1647, 1393-1652, 1395-1597, 1481-2145, 1542-1783, 1697-1785, 1781-2292, 1787-2082, 1793-2362, 1793-2487, 1793-2552, 1793-2568, 1795-2385, 1795-2463, 1796-2495, 1797-2322, 1804-2299, 1860-1989, 1860-2491, 1912-2330, 2345-2940, 2346-2997, 2403-2979, 2404-2633, 2405-2633, 2468-2957, 2504-2633, 2509-2994, 2825-2967, 2928-3507, 2979-3523, 2989-3489, 2999-3222, 2999-3384, 2999-3443, 2999-3491, 2999-3494, 2999-3502, 2999-3536, 2999-3551, 2999-
	3563, 2999-3669, 2999-3692, 2999-3799, 3000-3178, 3002-3249, 3122-3917, 3136-3740, 3142-3914, 3181-3597, 3195-3780, 3197-3917, 3221-3473, 3223-3917, 3259-3917, 3263-3917, 3277-3917, 3331-3917, 3354-3826, 3359-3631, 3359-3913, 3359-3917, 3364-3920, 3369-3714, 3451-3917, 3456-4156, 3456-4236, 3458-3917, 3479-3897, 3483-3925, 3485-3897, 3494-3897, 3495-3925, 3502-3917, 3502-3923, 3506-3897, 3514-3897, 3514-3917, 3525-3917, 3527-3917, 3538-3907, 3543-3917, 3550-3917, 3553-3917, 3561-3916, 3579-3917, 3601-3925, 3607-3925, 3609-3867, 3683-3917, 3706-4381, 3721-4394, 3751-3925, 3764-3921, 3918-4300, 3918-4358, 3918-4398, 3919-4385, 3924-4524, 3926-4367, 3926-4385, 3926-4398, 3926-4401, 3926-4486, 3936-4212, 3971-4205, 3971-4284, 3985-4611, 4040-4681, 4060-4737, 4116-4389, 4118-4703, 4128-4458, 4130-4716, 4155-4455, 4160-4407, 4179-4736, 4195-4663, 4223-4728, 4286-4902, 4290-4538, 4292-5015, 4341-4810, 4353-4999, 4371-4883, 4400-4657, 4411-4985, 4493-5130, 4511-5314, 4521-4760, 4521-4981, 4521-4988, 4557-5111, 4557-5207, 4567-5033, 4567-5034, 4617-4900, 4674-5203, 4681-5077, 4687-4933, 4687-4937, 4687-4950, 4706-4993, 4715-5153, 4729-5209,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
23 (cont.)	4731-5098, 4742-5322, 4770-5346, 4808-5046, 4808-5256, 4823-5309, 4825-5094, 4868-5535, 4894-5088, 4894-5377, 4895-5209, 4898-5520, 4904-5198, 4906-5520, 4909-5484, 4918-5533, 4947-5488, 4949-5601, 4964-5454, 4972-5156, 4978-5156, 4982-5603, 4988-5623, 4998-5601, 4999-5465, 5044-5542, 5049-5480, 5050-5395, 5059-5571, 5059-5584, 5077-5650, 5106-5378, 5134-5358, 5136-5400, 5151-5586, 5152-5429, 5153-5437, 5153-5644, 5153-5762, 5195-5480, 5241-5699, 5241-5702, 5247-5551, 5248-5821, 5268-5886, 5278-5534, 5279-5888, 5305-5970, 5313-5969, 5317-5602, 5358-5989, 5358-6062, 5364-5516, 5366-5639, 5378-5867, 5382-5600, 5382-5662, 5407-6033, 5417-5710, 5418-5759, 5419-5603, 5419-5671, 5419-5686, 5421-5629, 5422-5617, 5422-5791, 5422-5854, 5426-5682, 5432-6082, 5435-6109, 5481-5603, 5493-5739, 5493-6158, 5509-6104, 5511-5958, 5526-5794, 5537-5804, 5537-6124, 5538-6123, 5553-5868, 5556-5743, 5558-5849, 5559-5798, 5563-5814, 5572-5837, 5578-6232, 5578-6256, 5580-5838, 5602-6265, 5616-5885, 5633-6019, 5644-5939, 5648-5973, 5648-6267, 5660-6232, 5671-6140, 5671-6311, 5681-5961, 5696-5978, 5703-5971, 5779-5922, 5779-6309, 5781-6015, 5791-6375, 5794-
	6378, 5796-6402, 5798-6263, 5811-6093, 5812-6058, 5813-6057, 5814-6050, 5824-6087, 5824-6097, 5827-6077, 5828-6125, 5838-6443, 5839-6342, 5852-6134, 5852-6284, 5855-6064, 5855-6413, 5857-6443, 5858-6062, 5858-6077, 5867-6278, 5869-6121, 5869-6251, 5875-6147, 5875-6479, 5878-6051, 5878-6142, 5878-6155, 5878-6502, 5884-6156, 5885-6167, 5886-6041, 5887-6317, 5903-6080, 5903-6153, 5906-6094, 5907-6138, 5912-6140, 5916-6170, 5923-6506, 5932-6172, 5932-6213, 5940-6203, 5949-6364, 5949-6391, 5961-6199, 5970-6575, 5972-6182, 5983-6392, 5991-6219, 6008-6217, 6008-6238, 6008-6290, 6008-6295, 6009-6254, 6009-6337, 6009-6541, 6011-6253, 6011-6317, 6011-6532, 6011-6597, 6029-6345, 6029-6622, 6036-6276, 6036-6329, 6047-6776, 6050-6287, 6050-6289, 6051-6733, 6052-6125, 6075-6666, 6076-6571, 6076-6740, 6077-6714, 6087-6322, 6087-6355, 6087-6377, 6087-6484, 6103-6794, 6109-6373, 6110-6163, 6111-6331, 6111-6368, 6111-6373, 6112-6377, 6113-6372, 6114-6342, 6114-6376, 6114-6388, 6118-6340, 6120-6626, 6128-6338, 6128-6528, 6128-6617, 6133-6802, 6136-6685, 6140-6338, 6145-6182, 6145-6788, 6165-6413, 6172-6426, 6174-6678, 6178-6788, 6183-6657, 6183-6809,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
23 (cont.)	6212-6812, 6217-6472, 6219-6463, 6219-6744, 6228-6808, 6229-6457, 6232-6809, 6234-6812, 6239-6758, 6239-6774, 6244-6506, 6252-6462, 6252-6555, 6258-6541, 6258-6812, 6263-6510, 6264-6536, 6266-6551, 6271-6713, 6271-6762, 6272-6560, 6276-6812, 6281-6812, 6284-6530, 6291-6548, 6304-6803, 6324-6782, 6329-6802, 6336-6591, 6347-6617, 6347-6812, 6352-6802, 6353-6628, 6354-6527, 6354-6804, 6354-6812, 6355-6812, 6356-6800, 6358-6804, 6359-6802, 6366-6621, 6366-6808, 6367-6812, 6371-6812, 6373-6812, 6377-6599, 6380-6619, 6380-6631, 6380-6792, 6380-6812, 6385-6802, 6386-6804, 6387-6802, 6391-6803, 6394-6802, 6399-6807, 6403-6804, 6409-6807, 6409-6812, 6410-6489, 6411-6808, 6411-6812, 6414-6804, 6418-6806, 6418-6809, 6419-6802, 6425-6805, 6426-6812, 6427-6802, 6431-6812, 6438-6802, 6438-6804, 6440-6800, 6442-6804, 6443-6704, 6454-6781, 6459-6812, 6463-6802, 6469-6802, 6471-6812, 6472-6802, 6472-6804, 6476-6812, 6494-6805, 6497-6800, 6499-6802, 6501-6679, 6505-6804, 6507-6802, 6507-6812, 6509-6805, 6510-6803, 6513-6808, 6517-6800, 6518-6749, 6520-6802, 6525-6802, 6535-6782, 6547-6802, 6547-6812, 6564-6789, 6571-6804, 6572-6729, 6580-6804, 6586-6802, 6610-6811, 6616-6802, 6627-6802, 6628-6732, 6628-6763, 6628-6805, 6673-6762
24/7503618CB1/1589	1-1295, 1-1589, 48-284, 49-147, 49-166, 49-185, 49-196, 49-201, 49-215, 49-218, 49-227, 49-228, 49-231, 49-232, 49-234, 49-235, 49-237, 49-239, 49-240, 49-244, 49-247, 49-248, 49-252, 49-253, 49-255, 49-256, 49-258, 49-261, 49-263, 49-265, 49-270, 49-273, 49-277, 49-278, 49-279, 49-280, 49-283, 49-284, 49-286, 49-288, 49-289, 49-291, 49-292, 49-295, 49-296, 49-298, 49-303, 49-305, 49-311, 49-314, 49-333, 49-335, 49-345, 49-379, 49-423, 49-446, 49-449, 49-551, 49-612, 49-616, 49-641, 49-649, 50-246, 56-304, 59-336, 62-344, 64-423, 69-254, 70-412, 71-490, 72-332, 73-349, 80-650, 83-358, 103-361, 103-528, 105-574, 106-380, 107-505, 112-253, 117-566, 119-517, 119-584, 122-603, 124-448, 130-423, 130-549, 133-338, 135-230, 138-511, 148-362, 154-374, 164-623, 171-417, 172-371, 174-435, 177-285, 179-404, 179-441, 179-450, 182-333, 183-385, 183-389, 190-420, 197-490, 200-389, 200-495, 209-440, 209-475, 209-613, 213-493, 214-379, 224-509, 225-469, 230-467, 234-458, 236-438, 242-437, 244-412, 245-476, 251-550, 253-474, 253-479, 255-545, 258-481, 258-484, 258-650, 259-453, 259-569, 262-374, 263-587, 265-499, 270-485, 275-505, 285-515, 288-551, 290-443, 290-514, 290-540, 295-542, 297-514, 299-520, 300-

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
24 (cont.)	539, 300-576, 308-503, 316-609, 317-549, 317-580, 323-545, 329-543, 330-585, 337-613, 342-597, 342-612, 343-542, 344-484, 344-603, 346-649, 349-615, 354-449, 365-649, 383-641, 390-602, 391-634, 393-615, 397-650, 407-603, 419-650, 421-511, 421-617, 426-625, 429-650, 433-608, 433-650, 622-819, 647-858, 647-875, 647-882, 647-901, 652-922, 653-862, 658-889, 661-892, 661-1221, 666-1204, 668-932, 671-911, 671-941, 673-918, 676-934, 678-896, 678-905, 679-903, 679-915, 681-929, 696-904, 696-905, 696-914, 696-946, 696-960, 708-937, 710-1062, 714-1053, 718-1128, 719-953, 727-931, 727-975, 727-990, 727-1017, 728-971, 728-990, 728-1002, 729-1167, 733-949, 733-992, 744-978, 749-1026, 760-989, 760-994, 763-1048, 770-999, 770-1002, 772-971, 780-1053, 781-1035, 787-939, 788-1021, 794-1198, 796-920, 796-950, 796-979, 797-1065, 798-1069, 806-1021, 808-1010, 809-1010, 811-1051, 811-1069, 816-1229, 819-1075, 821-1084, 823-1191, 831-1078, 848-1161, 859-1127, 860-1057, 861-1096, 861-1118, 862-1131, 866-1087, 871-1151, 873-1165, 878-1116, 881-1206, 881-1243, 881-1279, 885-1170, 885-1178, 890-1290, 891-1140, 895-1122, 899-1137, 900-1246, 920-1102, 921-1241, 927-1185, 950-1200, 974-1161,
	979-1235, 981-1138, 984-1174, 1003-1258, 1010-1201, 1011-1134, 1011-1243, 1011-1297, 1012-1281, 1013-1234, 1014-1297, 1015-1214, 1018-1235, 1019-1259, 1031-1280, 1037-1294, 1044-1273, 1049-1177, 1061-1188, 1061-1297, 1063-1282, 1063-1296, 1064-1293, 1066-1297, 1070-1270, 1070-1277, 1082-1297, 1394-1589

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
13	1567742CB1	NGANNOT01
14	7485501CB1	SPLNNOT04
15	3089944CB1	SKINBIT01
16	5284076CB1	TESTNON04
17	2899903CB1	BRABDIE02
18	7491355CB1	PROSTUT09
19	3333288CB1	BRAIFER06
20	7488313CB1	COLNNOT01
21	6013113CB1	BRATNOT05
22	7488573CB1	OVARDIR01
23	7506027CB1	BRABDIE02
24	7503618CB1	CARGDIT01

Table 6

Library	Vector	Library Description
BRABDIE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day, for 40 years).
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRATNOT05	pINCY	Library was constructed using RNA isolated from temporal cortex tissue removed from a 45-year-old Caucasian female who died from a dissecting aortic aneurysm and ischemic bowel disease. Pathology indicated mild arteriosclerosis involving the cerebral cortical white matter and basal ganglia. Grossly, there was mild meningeal fibrosis and mild focal atherosclerotic plaque in the middle cerebral artery, as well as vertebral arteries bilaterally. Microscopically, the cerebral hemispheres, brain stem and cerebellum reveal focal areas in the white matter that contain blood vessels that were barrel-shaped, hyalinized, with hemosiderin-laden macrophages in the Virchow-Robin space. In addition, there were scattered neurofibrillary tangles within the basolateral nuclei of the amygdala. Patient history included mild atheromatosis of aorta and coronary arteries, bowel and liver infarct due to aneurysm, physiologic fatty liver associated with obesity, mild diffuse emphysema, thrombosis of mesenteric and portal veins, cardiomegaly due to hypertrophy of left ventricle, arterial hypertension, acute pulmonary edema, splenomegaly, obesity (300 lb.), leiomyoma of uterus, sleep apnea, and iron deficiency anemia.
CARGDIT01	pINCY	Library was constructed using RNA isolated from diseased cartilage tissue. Patient history included osteoarthritis.
COLNNOT01	PSPORT1	Library was constructed using RNA isolated from colon tissue removed from a 75-year-old Caucasian male during a hemicolectomy.
NGANNOT01	PSPORT1	Library was constructed using RNA isolated from tumorous neuroganglion tissue removed from a 9-year-old Caucasian male during a soft tissue excision of the chest wall. Pathology indicated a ganglioneuroma. Family history included asthma.

Table 6

Library	Vector	Library Description
OVARDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SPLNNOT04	pINCY	Library was constructed using RNA isolated from the spleen tissue of a 2-year-old Hispanic male, who died from cerebral anoxia. Past medical history and serologies were negative.
TESTNON04	pINCY	This normalized testis tissue library was constructed from 6.48 million independent clones from a pool of testis tissue libraries. Starting RNA was made from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al.,Genome Research 6 (1996):791, except that a significantly longer (48-hours/round)reannealing hybridization was used.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM and SMART.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM or SMART hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,
b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 and SEQ ID NO:12,
c) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:11,
d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and
e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:
a) culturing a cell under conditions suitable for expression of the polypeptide, wherein

said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,
b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-23,
c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:24,
d) a polynucleotide complementary to a polynucleotide of a),
e) a polynucleotide complementary to a polynucleotide of b),
f) a polynucleotide complementary to a polynucleotide of c), and
g) an RNA equivalent of a)-f).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

19. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of

claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of CGDD in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
b) isolating antibodies from said animal, and
c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
b) isolating antibody producing cells from the animal,
c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
d) culturing the hybridoma cells, and
e) isolating from the culture monoclonal antibody which binds specifically to a

polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

<110> INCYTE GENOMICS, INC.
YUE, Henry
YAO, Monique G.
ISON, Craig H.
LU, Yan
WARREN, Bridget A.
ELLIOTT, Vicki S.
BAUGHN, Mariah R.
DING, Li
XU, Yuming
GIETZEN, Kimberly J.
TANG, Tom Y.
LAL, Preeti
DUGGAN, Brendan M.
BURFORD, Neil
LU, Dyung Aina M.
RICHARDSON, Thomas W.
TRAN, Uyen K.
KHARE, Reena
WALIA, Narinder K.

<120> PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

<130> PF-0903 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/268,111; 60/271,175; 60/274,503; 60/274,552
<151> 2001-02-09; 2001-02-23; 2001-03-08; 2001-03-09

<160> 24
<170> PERL Program

<210> 1
<211> 977
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1567742CD1

<400> 1
```
Met Ala Ser Ser His Ser Ser Ser Pro Val Pro Gln Gly Ser Ser
  1               5                  10                  15
Ser Asp Val Phe Phe Lys Ile Glu Val Asp Pro Ser Lys His Ile
                 20                  25                  30
Arg Pro Val Pro Ser Leu Pro Asp Val Cys Pro Lys Glu Pro Thr
                 35                  40                  45
Gly Asp Ser His Ser Leu Tyr Val Ala Pro Ser Leu Val Thr Asp
                 50                  55                  60
Gln His Arg Trp Thr Val Tyr His Ser Lys Val Asn Leu Pro Ala
                 65                  70                  75
Ala Leu Asn Asp Pro Arg Leu Ala Lys Arg Glu Ser Asp Phe Phe
                 80                  85                  90
```

```
Thr Lys Thr Trp Gly Leu Asp Phe Val Asp Thr Glu Val Ile Pro
                95                  100                 105
Ser Phe Tyr Leu Pro Gln Ile Ser Lys Glu His Phe Thr Val Tyr
                110                 115                 120
Gln Gln Glu Ile Ser Gln Arg Glu Lys Ile His Glu Arg Cys Lys
                125                 130                 135
Asn Ile Cys Pro Pro Lys Asp Thr Phe Glu Arg Thr Leu Leu His
                140                 145                 150
Thr His Asp Lys Ser Arg Thr Asp Leu Glu Gln Val Pro Lys Ile
                155                 160                 165
Phe Met Lys Pro Asp Phe Ala Leu Asp Asp Ser Leu Thr Phe Asn
                170                 175                 180
Ser Val Leu Pro Trp Ser His Phe Asn Thr Ala Gly Gly Lys Gly
                185                 190                 195
Asn Arg Asp Ala Ala Ser Ser Lys Leu Leu Gln Glu Lys Leu Ser
                200                 205                 210
His Tyr Leu Asp Ile Val Glu Val Asn Ile Ala His Gln Ile Ser
                215                 220                 225
Leu Arg Ser Glu Ala Phe Phe His Ala Met Thr Ser Gln His Glu
                230                 235                 240
Leu Gln Asp Tyr Leu Arg Lys Thr Ser Gln Ala Val Lys Met Leu
                245                 250                 255
Arg Asp Lys Ile Ala Gln Ile Asp Lys Val Met Cys Glu Gly Ser
                260                 265                 270
Leu His Ile Leu Arg Leu Ala Leu Thr Arg Asn Asn Cys Val Lys
                275                 280                 285
Val Tyr Asn Lys Leu Lys Leu Met Ala Thr Val His Gln Thr Gln
                290                 295                 300
Pro Thr Val Gln Val Leu Leu Ser Thr Ser Glu Phe Val Gly Ala
                305                 310                 315
Leu Asp Leu Ile Ala Thr Thr Gln Glu Val Leu Gln Gln Glu Leu
                320                 325                 330
Gln Gly Ile His Ser Phe Arg His Leu Gly Ser Gln Leu Cys Glu
                335                 340                 345
Leu Glu Lys Leu Ile Asp Lys Met Met Ile Ala Glu Phe Ser Thr
                350                 355                 360
Tyr Ser His Ser Asp Leu Asn Arg Pro Leu Glu Asp Asp Cys Gln
                365                 370                 375
Val Leu Glu Glu Glu Arg Leu Ile Ser Leu Val Phe Gly Leu Leu
                380                 385                 390
Lys Gln Arg Lys Leu Asn Phe Leu Glu Ile Tyr Gly Glu Lys Met
                395                 400                 405
Val Ile Thr Ala Lys Asn Ile Ile Lys Gln Cys Val Ile Asn Lys
                410                 415                 420
Val Ser Gln Thr Glu Glu Ile Asp Thr Asp Val Val Val Lys Leu
                425                 430                 435
Ala Asp Gln Met Arg Met Leu Asn Phe Pro Gln Trp Phe Asp Leu
                440                 445                 450
Leu Lys Asp Ile Phe Ser Lys Phe Thr Ile Phe Leu Gln Arg Val
                455                 460                 465
Lys Ala Thr Leu Asn Ile Ile His Ser Val Val Leu Ser Val Leu
                470                 475                 480
Asp Lys Asn Gln Arg Thr Arg Glu Leu Glu Glu Ile Ser Gln Gln
                485                 490                 495
Lys Asn Ala Ala Lys Asp Asn Ser Leu Asp Thr Glu Val Ala Tyr
                500                 505                 510
```

Leu Ile His Glu Gly Met Phe Ile Ser Asp Ala Phe Gly Glu Gly
515 520 525
Glu Leu Thr Pro Ile Ala Val Asp Thr Thr Ser Gln Arg Asn Ala
530 535 540
Ser Pro Asn Ser Glu Pro Cys Ser Ser Asp Ser Val Ser Glu Pro
545 550 555
Glu Cys Thr Thr Asp Ser Ser Ser Ser Lys Glu His Thr Ser Ser
560 565 570
Ser Ala Ile Pro Gly Gly Val Asp Ile Met Val Ser Glu Asp Met
575 580 585
Lys Leu Thr Asp Ser Glu Leu Gly Lys Leu Ala Asn Asn Ile Gln
590 595 600
Glu Leu Leu Tyr Ser Ala Ser Asp Ile Cys His Asp Arg Ala Val
605 610 615
Lys Phe Leu Met Ser Arg Ala Lys Asp Gly Phe Leu Glu Lys Leu
620 625 630
Asn Ser Met Glu Phe Ile Thr Leu Ser Arg Leu Met Glu Thr Phe
635 640 645
Ile Leu Asp Thr Glu Gln Ile Cys Gly Arg Lys Ser Thr Ser Leu
650 655 660
Leu Gly Ala Leu Gln Ser Gln Ala Ile Lys Phe Val Asn Arg Phe
665 670 675
His Glu Glu Arg Lys Thr Lys Leu Ser Leu Leu Leu Asp Asn Glu
680 685 690
Arg Trp Lys Gln Ala Asp Val Pro Ala Glu Phe Gln Asp Leu Val
695 700 705
Asp Ser Leu Ser Asp Gly Lys Ile Ala Leu Pro Glu Lys Lys Ser
710 715 720
Gly Ala Thr Glu Glu Arg Lys Pro Ala Glu Val Leu Ile Val Glu
725 730 735
Gly Gln Gln Tyr Ala Val Val Gly Thr Val Leu Leu Leu Ile Arg
740 745 750
Ile Ile Leu Glu Tyr Cys Gln Cys Val Asp Asn Ile Pro Ser Val
755 760 765
Thr Thr Asp Met Leu Thr Arg Leu Ser Asp Leu Leu Lys Tyr Phe
770 775 780
Asn Ser Arg Ser Cys Gln Leu Val Leu Gly Ala Gly Ala Leu Gln
785 790 795
Val Val Gly Leu Lys Thr Ile Thr Thr Lys Asn Leu Ala Leu Ser
800 805 810
Ser Arg Cys Leu Gln Leu Ile Val His Tyr Ile Pro Val Ile Arg
815 820 825
Ala His Phe Glu Ala Arg Leu Pro Pro Lys Gln Tyr Ser Met Leu
830 835 840
Arg His Phe Asp His Ile Thr Lys Asp Tyr His Asp His Ile Ala
845 850 855
Glu Ile Ser Ala Lys Leu Val Ala Ile Met Asp Ser Leu Phe Asp
860 865 870
Lys Leu Leu Ser Lys Tyr Glu Val Lys Ala Pro Val Pro Ser Ala
875 880 885
Cys Phe Arg Asn Ile Cys Lys Gln Met Thr Lys Met His Glu Ala
890 895 900
Ile Phe Asp Leu Leu Pro Glu Glu Gln Thr Gln Met Leu Phe Leu
905 910 915
Arg Ile Asn Ala Ser Tyr Lys Leu His Leu Lys Lys Gln Leu Ser
920 925 930

His Leu Asn Val Ile Asn Asp Gly Gly Pro Gln Asn Gly Leu Val
935 940 945
Thr Ala Asp Val Ala Phe Tyr Thr Gly Asn Leu Gln Ala Leu Lys
950 955 960
Gly Leu Lys Asp Leu Asp Leu Asn Met Ala Glu Ile Trp Glu Gln
965 970 975

Lys Arg

<210> 2
<211> 109
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485501CD1

<400> 2
Met Ser Tyr Lys Pro Thr Thr Pro Ala Pro Ser Ser Thr Pro Gly
1 5 10 15
Phe Ser Thr Pro Gly Pro Gly Thr Pro Val Pro Thr Gly Ser Val
20 25 30
Pro Ser Pro Ser Gly Ser Gly Pro Gly Ala Thr Ala Pro Cys Arg
35 40 45
Pro Leu Phe Lys Asp Phe Gly Pro Pro Thr Val Gly Cys Val Gln
50 55 60
Ala Met Lys Pro Pro Gly Ala Gln Gly Ser Gln Ser Thr Tyr Thr
65 70 75
Glu Leu Leu Leu Val Thr Gly Glu Met Gly Lys Gly Ile Arg Pro
80 85 90
Thr Tyr Ala Gly Ser Lys Ser Ala Ala Glu Arg Leu Lys Arg Gly
95 100 105
Ile Ile His Pro

<210> 3
<211> 468
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3089944CD1

<400> 3
Met Ala Ala Pro Asp Leu Ser Thr Asn Leu Gln Glu Glu Ala Thr
1 5 10 15
Cys Ala Ile Cys Leu Asp Tyr Phe Thr Asp Pro Val Met Thr Asp
20 25 30
Cys Gly His Asn Phe Cys Arg Glu Cys Ile Arg Arg Cys Trp Gly
35 40 45
Gln Pro Glu Gly Pro Tyr Ala Cys Pro Glu Cys Arg Glu Leu Ser
50 55 60
Pro Gln Arg Asn Leu Arg Pro Asn Arg Pro Leu Ala Lys Met Ala
65 70 75

Glu Met Ala Arg Arg Leu His Pro Pro Ser Pro Val Pro Gln Gly
80 85 90

Val Cys Pro Ala His Arg Glu Pro Leu Ala Ala Phe Cys Gly Asp
95 100 105

Glu Leu Arg Leu Leu Cys Ala Ala Cys Glu Arg Ser Gly Glu His
110 115 120

Trp Ala His Arg Val Arg Pro Leu Gln Asp Ala Ala Glu Asp Leu
125 130 135

Lys Ala Lys Leu Glu Lys Ser Leu Glu His Leu Arg Lys Gln Met
140 145 150

Gln Asp Ala Leu Leu Phe Gln Ala Gln Ala Asp Glu Thr Cys Val
155 160 165

Leu Trp Gln Lys Met Val Glu Ser Gln Arg Gln Asn Val Leu Gly
170 175 180

Glu Phe Glu Arg Leu Arg Arg Leu Leu Ala Glu Glu Glu Gln Gln
185 190 195

Leu Leu Gln Arg Leu Glu Glu Glu Glu Leu Glu Val Leu Pro Arg
200 205 210

Leu Arg Glu Gly Ala Ala His Leu Gly Gln Gln Ser Ala His Leu
215 220 225

Ala Glu Leu Ile Ala Glu Leu Glu Gly Arg Cys Gln Leu Pro Ala
230 235 240

Leu Gly Leu Leu Gln Asp Ile Lys Asp Ala Leu Arg Arg Val Gln
245 250 255

Asp Val Lys Leu Gln Pro Pro Glu Val Val Pro Met Glu Leu Arg
260 265 270

Thr Val Cys Arg Val Pro Gly Leu Val Glu Thr Leu Arg Arg Phe
275 280 285

Arg Gly Asp Val Thr Leu Asp Pro Asp Thr Ala Asn Pro Glu Leu
290 295 300

Ile Leu Ser Glu Asp Arg Arg Ser Val Gln Arg Gly Asp Leu Arg
305 310 315

Gln Ala Leu Pro Asp Ser Pro Glu Arg Phe Asp Pro Gly Pro Cys
320 325 330

Val Leu Gly Gln Glu Arg Phe Thr Ser Gly Arg His Tyr Trp Glu
335 340 345

Val Glu Val Gly Asp Arg Thr Ser Trp Ala Leu Gly Val Cys Arg
350 355 360

Glu Asn Val Asn Arg Lys Glu Lys Gly Glu Leu Ser Ala Gly Asn
365 370 375

Gly Phe Trp Ile Leu Val Phe Leu Gly Ser Tyr Tyr Asn Ser Ser
380 385 390

Glu Arg Ala Leu Ala Pro Leu Arg Asp Pro Pro Arg Arg Val Gly
395 400 405

Ile Phe Leu Asp Tyr Glu Ala Gly His Leu Ser Phe Tyr Ser Ala
410 415 420

Thr Asp Gly Ser Leu Leu Phe Ile Phe Pro Glu Ile Pro Phe Ser
425 430 435

Gly Thr Leu Arg Pro Leu Phe Ser Pro Leu Ser Ser Ser Pro Thr
440 445 450

Pro Met Thr Ile Cys Arg Pro Lys Gly Gly Ser Gly Asp Thr Leu
455 460 465

Ala Pro Gln

<210> 4

<211> 158
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5284076CD1

<400> 4

```
Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile Trp Thr
  1               5                  10                  15
Gly Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp Trp
                 20                  25                  30
Leu Met Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr
                 35                  40                  45
Ile Phe Ala Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn
                 50                  55                  60
Arg Ile His Thr Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp
                 65                  70                  75
Cys Gly Ile Arg Thr Arg Val Val Ser Glu Glu Thr Leu Leu Phe
                 80                  85                  90
Gln Thr Glu Leu Tyr Phe Thr Pro Arg Asn Ile Asp His Asp Pro
                 95                 100                 105
Gln Glu Ile His Leu Glu Cys Ser Thr Ser Arg Lys Ser Val Trp
                110                 115                 120
Leu Thr Pro Val Ser Thr Glu Asn Glu Ile Lys Leu Asp Pro Ser
                125                 130                 135
Pro Phe Ile Ala Asp Phe Gln Thr Thr Ala Glu Glu Leu Gly Leu
                140                 145                 150
Leu Ser Ser Ser Pro Asn Leu Leu
                155
```

<210> 5
<211> 1161
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2899903CD1

<400> 5

```
Met Glu Pro Asp Ser Leu Leu Asp Gln Asp Asp Ser Tyr Glu Ser
  1               5                  10                  15
Pro Gln Glu Arg Pro Gly Ser Arg Arg Ser Leu Pro Gly Ser Leu
                 20                  25                  30
Ser Glu Lys Ser Pro Ser Met Glu Pro Ser Ala Ala Thr Pro Phe
                 35                  40                  45
Arg Val Thr Gly Phe Leu Ser Arg Arg Leu Lys Gly Ser Ile Lys
                 50                  55                  60
Arg Thr Lys Ser Gln Pro Lys Leu Asp Arg Asn His Ser Phe Arg
                 65                  70                  75
His Ile Leu Pro Gly Phe Arg Ser Ala Ala Ala Ala Ala Ala Asp
                 80                  85                  90
Asn Glu Arg Ser His Leu Met Pro Arg Leu Lys Glu Ser Arg Ser
                 95                 100                 105
```

```
His Glu Ser Leu Leu Ser Pro Ser Ser Ala Val Glu Ala Leu Asp
                110                 115                 120
Leu Ser Met Glu Glu Glu Val Val Ile Lys Pro Val His Ser Ser
                125                 130                 135
Ile Leu Gly Gln Asp Tyr Cys Phe Glu Val Thr Thr Ser Ser Gly
                140                 145                 150
Ser Lys Cys Phe Ser Cys Arg Ser Ala Ala Glu Arg Asp Lys Trp
                155                 160                 165
Met Glu Asn Leu Arg Arg Ala Val His Pro Asn Lys Asp Asn Ser
                170                 175                 180
Arg Arg Val Glu His Ile Leu Lys Leu Trp Val Ile Glu Ala Lys
                185                 190                 195
Asp Leu Pro Ala Lys Lys Lys Tyr Leu Cys Glu Leu Cys Leu Asp
                200                 205                 210
Asp Val Leu Tyr Ala Arg Thr Thr Gly Lys Leu Lys Thr Asp Asn
                215                 220                 225
Val Phe Trp Gly Glu His Phe Glu Phe His Asn Leu Pro Pro Leu
                230                 235                 240
Arg Thr Val Thr Val His Leu Tyr Arg Glu Thr Asp Lys Lys Lys
                245                 250                 255
Lys Lys Glu Arg Asn Ser Tyr Leu Gly Leu Val Ser Leu Pro Ala
                260                 265                 270
Ala Ser Val Ala Gly Arg Gln Phe Val Glu Lys Trp Tyr Pro Val
                275                 280                 285
Val Thr Pro Asn Pro Lys Gly Gly Lys Gly Pro Gly Pro Met Ile
                290                 295                 300
Arg Ile Lys Ala Arg Tyr Gln Thr Ile Thr Ile Leu Pro Met Glu
                305                 310                 315
Met Tyr Lys Glu Phe Ala Glu His Ile Thr Asn His Tyr Leu Gly
                320                 325                 330
Leu Cys Ala Ala Leu Glu Pro Ile Leu Ser Ala Lys Thr Lys Glu
                335                 340                 345
Glu Met Ala Ser Ala Leu Val His Ile Leu Gln Ser Thr Gly Lys
                350                 355                 360
Val Lys Asp Phe Leu Thr Asp Leu Met Met Ser Glu Val Asp Arg
                365                 370                 375
Cys Gly Asp Asn Glu His Leu Ile Phe Arg Glu Asn Thr Leu Ala
                380                 385                 390
Thr Lys Ala Ile Glu Glu Tyr Leu Lys Leu Val Gly Gln Lys Tyr
                395                 400                 405
Leu Gln Asp Ala Leu Gly Glu Phe Ile Lys Ala Leu Tyr Glu Ser
                410                 415                 420
Asp Glu Asn Cys Glu Val Asp Pro Ser Lys Cys Ser Ala Ala Asp
                425                 430                 435
Leu Pro Glu His Gln Gly Asn Leu Lys Met Cys Cys Glu Leu Ala
                440                 445                 450
Phe Cys Lys Ile Ile Asn Ser Tyr Cys Val Phe Pro Arg Glu Leu
                455                 460                 465
Lys Glu Val Phe Ala Ser Trp Arg Gln Glu Cys Ser Ser Arg Gly
                470                 475                 480
Arg Pro Asp Ile Ser Glu Arg Leu Ile Ser Ala Ser Leu Phe Leu
                485                 490                 495
Arg Phe Leu Cys Pro Ala Ile Met Ser Pro Ser Leu Phe Asn Leu
                500                 505                 510
Leu Gln Glu Tyr Pro Asp Asp Arg Thr Ala Arg Thr Leu Thr Leu
                515                 520                 525
```

Ile Ala Lys Val Thr Gln Asn Leu Ala Asn Phe Ala Lys Phe Gly
530 535 540
Ser Lys Glu Glu Tyr Met Ser Phe Met Asn Gln Phe Leu Glu His
545 550 555
Glu Trp Thr Asn Met Gln Arg Phe Leu Leu Glu Ile Ser Asn Pro
560 565 570
Glu Thr Leu Ser Asn Thr Ala Gly Phe Glu Gly Tyr Ile Asp Leu
575 580 585
Gly Arg Glu Leu Ser Ser Leu His Ser Leu Leu Trp Glu Ala Val
590 595 600
Ser Gln Leu Glu Gln Ser Ile Val Ser Lys Leu Gly Pro Leu Pro
605 610 615
Arg Ile Leu Arg Asp Val His Thr Ala Leu Ser Thr Pro Gly Ser
620 625 630
Gly Gln Leu Pro Gly Thr Asn Asp Leu Ala Ser Thr Pro Gly Ser
635 640 645
Gly Ser Ser Ser Ile Ser Ala Gly Leu Gln Lys Met Val Ile Glu
650 655 660
Asn Asp Leu Ser Gly Leu Ile Asp Phe Thr Arg Leu Pro Ser Pro
665 670 675
Thr Pro Glu Asn Lys Asp Leu Phe Phe Val Thr Arg Ser Ser Gly
680 685 690
Val Gln Pro Ser Pro Ala Arg Ser Ser Ser Tyr Ser Glu Ala Asn
695 700 705
Glu Pro Asp Leu Gln Met Ala Asn Gly Gly Lys Ser Leu Ser Met
710 715 720
Val Asp Leu Gln Asp Ala Arg Thr Leu Asp Gly Glu Ala Gly Ser
725 730 735
Pro Ala Gly Pro Asp Val Leu Pro Thr Asp Gly Gln Ala Ala Ala
740 745 750
Ala Gln Leu Val Ala Gly Trp Pro Ala Arg Ala Thr Pro Val Asn
755 760 765
Leu Ala Gly Leu Ala Thr Val Arg Arg Ala Gly Gln Thr Pro Thr
770 775 780
Thr Pro Gly Thr Ser Glu Gly Ala Pro Gly Arg Pro Gln Leu Leu
785 790 795
Ala Pro Leu Ser Phe Gln Asn Pro Val Tyr Gln Met Ala Ala Gly
800 805 810
Leu Pro Leu Ser Pro Arg Gly Leu Gly Asp Ser Gly Ser Glu Gly
815 820 825
His Ser Ser Leu Ser Ser His Ser Asn Ser Glu Glu Leu Ala Ala
830 835 840
Ala Ala Lys Leu Gly Ser Phe Ser Thr Ala Ala Glu Glu Leu Ala
845 850 855
Arg Arg Pro Gly Glu Leu Ala Arg Arg Gln Met Ser Leu Thr Glu
860 865 870
Lys Gly Gly Gln Pro Thr Val Pro Arg Gln Asn Ser Ala Gly Pro
875 880 885
Gln Arg Arg Ile Asp Gln Pro Pro Pro Pro Pro Pro Pro Pro Pro
890 895 900
Pro Ala Pro Arg Gly Arg Thr Pro Pro Asn Leu Leu Ser Thr Leu
905 910 915
Gln Tyr Pro Arg Pro Ser Ser Gly Thr Leu Ala Ser Ala Ser Pro
920 925 930
Asp Trp Val Gly Pro Ser Thr Arg Leu Arg Gln Gln Ser Ser Ser
935 940 945

Ser Lys Gly Asp Ser Pro Glu Leu Lys Pro Arg Ala Val His Lys
950 955 960
Gln Gly Pro Ser Pro Val Ser Pro Asn Ala Leu Asp Arg Thr Ala
965 970 975
Ala Trp Leu Leu Thr Met Asn Ala Gln Leu Leu Glu Asp Glu Gly
980 985 990
Leu Gly Pro Asp Pro Pro His Arg Asp Arg Leu Arg Ser Lys Asp
995 1000 1005
Glu Leu Ser Gln Ala Glu Lys Asp Leu Ala Val Leu Gln Asp Lys
1010 1015 1020
Leu Arg Ile Ser Thr Lys Lys Leu Glu Glu Tyr Glu Thr Leu Phe
1025 1030 1035
Lys Cys Gln Glu Glu Thr Thr Gln Lys Leu Val Leu Glu Tyr Gln
1040 1045 1050
Ala Arg Leu Glu Glu Gly Glu Glu Arg Leu Arg Arg Gln Gln Glu
1055 1060 1065
Asp Lys Asp Ile Gln Met Lys Gly Ile Ile Ser Arg Leu Met Ser
1070 1075 1080
Val Glu Glu Glu Leu Lys Lys Asp His Ala Glu Met Gln Ala Ala
1085 1090 1095
Val Asp Ser Lys Gln Lys Ile Ile Asp Ala Gln Glu Lys Arg Ile
1100 1105 1110
Ala Ser Leu Asp Ala Ala Asn Ala Arg Leu Met Ser Ala Leu Thr
1115 1120 1125
Gln Leu Lys Glu Arg Tyr Ser Met Gln Ala Arg Asn Gly Ile Ser
1130 1135 1140
Pro Thr Asn Pro Thr Lys Leu Gln Ile Thr Glu Asn Gly Glu Phe
1145 1150 1155
Arg Asn Ser Ser Asn Cys
1160

<210> 6
<211> 331
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491355CD1

<400> 6
Met Ser Arg Ala Arg Gly Ala Leu Cys Arg Ala Cys Leu Ala Leu
1 5 10 15
Ala Ala Ala Leu Ala Ala Leu Leu Leu Leu Pro Leu Pro Leu Pro
20 25 30
Arg Ala Pro Ala Pro Ala Arg Thr Pro Ala Pro Ala Pro Arg Ala
35 40 45
Pro Pro Ser Arg Pro Ala Ala Pro Ser Leu Arg Pro Asp Asp Val
50 55 60
Phe Ile Ala Val Lys Thr Thr Arg Lys Asn His Gly Pro Arg Leu
65 70 75
Leu Leu Leu Leu Arg Thr Trp Ile Ser Arg Ala Arg Gln Gln Thr
80 85 90
Phe Ile Phe Thr Asp Gly Asp Asp Pro Glu Leu Glu Leu Gln Gly
95 100 105
Gly Asp Arg Val Ile Asn Thr Asn Cys Ser Ala Val Arg Thr Arg

```
                    110                          115                          120
Gln Ala Leu Cys Cys Lys Met Ser Val Glu Tyr Asp Lys Phe Ile
                125                          130                          135
Glu Ser Gly Arg Lys Trp Phe Cys His Val Asp Asp Asp Asn Tyr
                140                          145                          150
Val Asn Ala Arg Ser Leu Leu His Leu Leu Ser Ser Phe Ser Pro
                155                          160                          165
Ser Gln Asp Val Tyr Leu Gly Arg Pro Ser Leu Asp His Pro Ile
                170                          175                          180
Glu Ala Thr Glu Arg Val Gln Gly Gly Arg Thr Val Thr Thr Val
                185                          190                          195
Lys Phe Trp Phe Ala Thr Gly Gly Ala Gly Phe Cys Leu Ser Arg
                200                          205                          210
Gly Leu Ala Leu Lys Met Ser Pro Trp Ala Ser Leu Gly Ser Phe
                215                          220                          225
Met Ser Thr Ala Glu Gln Val Arg Leu Pro Asp Asp Cys Thr Val
                230                          235                          240
Gly Tyr Ile Val Glu Gly Leu Leu Gly Ala Arg Leu Leu His Ser
                245                          250                          255
Pro Leu Phe His Ser His Leu Glu Asn Leu Gln Arg Leu Pro Pro
                260                          265                          270
Asp Thr Leu Leu Gln Gln Val Thr Leu Ser His Gly Gly Pro Glu
                275                          280                          285
Asn Pro Gln Asn Val Val Asn Val Ala Gly Gly Phe Ser Leu His
                290                          295                          300
Gln Asp Pro Thr Arg Phe Lys Ser Ile His Cys Leu Leu Tyr Pro
                305                          310                          315
Asp Thr Asp Trp Cys Pro Arg Gln Lys Gln Gly Ala Pro Thr Ser
                320                          325                          330
Arg
```

<210> 7
<211> 579
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3333288CD1

<400> 7

```
Met Ser Ala Leu Arg Pro Leu Leu Leu Leu Leu Leu Pro Leu Cys
  1               5                  10                  15
Pro Gly Pro Gly Pro Gly Pro Gly Ser Glu Ala Lys Val Thr Arg
                 20                  25                  30
Ser Cys Ala Glu Thr Arg Gln Val Leu Gly Ala Arg Gly Tyr Ser
                 35                  40                  45
Leu Asn Leu Ile Pro Pro Ala Leu Ile Ser Gly Glu His Leu Arg
                 50                  55                  60
Val Cys Pro Gln Glu Tyr Thr Cys Cys Ser Ser Glu Thr Glu Gln
                 65                  70                  75
Arg Leu Ile Arg Glu Thr Glu Ala Thr Phe Arg Gly Leu Val Glu
                 80                  85                  90
Asp Ser Gly Ser Phe Leu Val His Thr Leu Ala Ala Arg His Arg
                 95                 100                 105
```

Lys Phe Asp Glu Phe Phe Leu Glu Met Leu Ser Val Ala Gln His
110 115 120
Ser Leu Thr Gln Leu Phe Ser His Ser Tyr Gly Arg Leu Tyr Ala
125 130 135
Gln His Ala Leu Ile Phe Asn Gly Leu Phe Ser Arg Leu Arg Asp
140 145 150
Phe Tyr Gly Glu Ser Gly Glu Gly Leu Asp Asp Thr Leu Ala Asp
155 160 165
Phe Trp Ala Gln Leu Leu Glu Arg Val Phe Pro Leu Leu His Pro
170 175 180
Gln Tyr Ser Phe Pro Pro Asp Tyr Leu Leu Cys Leu Ser Arg Leu
185 190 195
Ala Ser Ser Thr Asp Gly Ser Leu Gln Pro Phe Gly Asp Ser Pro
200 205 210
Arg Arg Leu Arg Leu Gln Ile Thr Arg Thr Leu Val Ala Ala Arg
215 220 225
Ala Phe Val Gln Gly Leu Glu Thr Gly Arg Asn Val Val Ser Glu
230 235 240
Ala Leu Lys Val Pro Val Ser Glu Gly Cys Ser Gln Ala Leu Met
245 250 255
Arg Leu Ile Gly Cys Pro Leu Cys Arg Gly Val Pro Ser Leu Met
260 265 270
Pro Cys Gln Gly Phe Cys Leu Asn Val Val Arg Gly Cys Leu Ser
275 280 285
Ser Arg Gly Leu Glu Pro Asp Trp Gly Asn Tyr Leu Asp Gly Leu
290 295 300
Leu Ile Leu Ala Asp Lys Leu Gln Gly Pro Phe Ser Phe Glu Leu
305 310 315
Thr Ala Glu Ser Ile Gly Val Lys Ile Ser Glu Gly Leu Met Tyr
320 325 330
Leu Gln Glu Asn Ser Ala Lys Val Ser Ala Gln Val Phe Gln Glu
335 340 345
Cys Gly Pro Pro Asp Pro Val Pro Ala Arg Asn Arg Arg Ala Pro
350 355 360
Pro Pro Arg Glu Glu Ala Gly Arg Leu Trp Ser Met Val Thr Glu
365 370 375
Glu Glu Arg Pro Thr Thr Ala Ala Gly Thr Asn Leu His Arg Leu
380 385 390
Val Trp Glu Leu Arg Glu Arg Leu Ala Arg Met Arg Gly Phe Trp
395 400 405
Ala Arg Leu Ser Leu Thr Val Cys Gly Asp Ser Arg Met Ala Ala
410 415 420
Asp Ala Ser Leu Glu Ala Ala Pro Cys Trp Thr Gly Ala Gly Arg
425 430 435
Gly Arg Tyr Leu Pro Pro Val Val Gly Gly Ser Pro Ala Glu Gln
440 445 450
Val Asn Asn Pro Glu Leu Lys Val Asp Ala Ser Gly Pro Asp Val
455 460 465
Pro Thr Arg Arg Arg Arg Leu Gln Leu Arg Ala Ala Thr Ala Arg
470 475 480
Met Lys Thr Ala Ala Leu Gly His Asp Leu Asp Gly Gln Asp Ala
485 490 495
Asp Glu Asp Ala Ser Gly Ser Gly Gly Gly Gln Gln Tyr Ala Asp
500 505 510
Asp Trp Met Ala Gly Ala Val Ala Pro Pro Ala Arg Pro Pro Arg
515 520 525

```
Pro Pro Tyr Pro Pro Arg Arg Asp Gly Ser Gly Gly Lys Gly Gly
                530                 535                 540
Gly Gly Ser Ala Arg Tyr Asn Gln Gly Arg Ser Arg Ser Gly Gly
                545                 550                 555
Ala Ser Ile Gly Phe His Thr Gln Thr Ile Leu Ile Leu Ser Leu
                560                 565                 570
Ser Ala Leu Ala Leu Leu Gly Pro Arg
                575

<210> 8
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488313CD1

<400> 8
Met Glu Gly Gln Asp Glu Val Ser Ala Arg Glu Gln His Phe His
  1               5                  10                  15
Ser Gln Val Arg Glu Ser Thr Ile Cys Phe Leu Leu Phe Ala Ile
                 20                  25                  30
Leu Tyr Val Val Ser Tyr Phe Ile Ile Thr Arg Tyr Lys Arg Lys
                 35                  40                  45
Ser Asp Glu Gln Glu Asp Glu Asp Ala Ile Val Asn Arg Ile Ser
                 50                  55                  60
Leu Phe Leu Ser Thr Phe Thr Leu Ala Val Ser Ala Gly Ala Val
                 65                  70                  75
Leu Leu Leu Pro Phe Ser Ile Ile Ser Asn Glu Ile Leu Leu Ser
                 80                  85                  90
Phe Pro Gln Asn Tyr Tyr Ile Gln Trp Leu Asn Gly Ser Leu Ile
                 95                 100                 105
His Gly Leu Trp Asn Leu Ala Ser Leu Phe Ser Asn Leu Cys Leu
                110                 115                 120
Phe Val Leu Met Pro Phe Ala Phe Phe Phe Leu Glu Ser Glu Gly
                125                 130                 135
Phe Ala Gly Leu Lys Lys Gly Ile Arg Ala Arg Ile Leu Glu Thr
                140                 145                 150
Leu Val Met Leu Leu Leu Leu Ala Leu Leu Ile Leu Gly Ile Val
                155                 160                 165
Trp Val Ala Ser Ala Leu Ile Asp Asn Asp Ala Ala Ser Met Glu
                170                 175                 180
Ser Leu Tyr Asp Leu Trp Glu Phe Tyr Leu Pro Tyr Leu Tyr Ser
                185                 190                 195
Cys Ile Ser Leu Met Gly Cys Leu Leu Leu Leu Leu Cys Thr Pro
                200                 205                 210
Val Gly Leu Ser Arg Met Phe Thr Val Met Gly Gln Leu Leu Val
                215                 220                 225
Lys Pro Thr Ile Leu Glu Asp Leu Asp Glu Gln Ile Tyr Ile Ile
                230                 235                 240
Thr Leu Glu Glu Glu Ala Leu Gln Arg Arg Leu Asn Gly Leu Ser
                245                 250                 255
Ser Ser Val Glu Tyr Asn Ile Met Glu Leu Glu Gln Glu Leu Glu
                260                 265                 270
Asn Val Lys Thr Leu Lys Thr Lys Leu Glu Arg Arg Lys Lys Ala
```

275 280 285
Ser Ala Trp Glu Arg Asn Leu Val Tyr Pro Ala Val Met Val Leu
290 295 300
Leu Leu Ile Glu Thr Ser Ile Ser Val Leu Leu Val Ala Cys Asn
305 310 315
Ile Leu Cys Leu Leu Val Asp Glu Thr Ala Met Pro Lys Gly Thr
320 325 330
Arg Gly Pro Gly Ile Gly Asn Ala Ser Leu Ser Thr Phe Gly Phe
335 340 345
Val Gly Ala Ala Leu Glu Ile Ile Leu Ile Phe Tyr Leu Met Val
350 355 360
Ser Ser Val Val Gly Phe Tyr Ser Leu Arg Phe Phe Gly Asn Phe
365 370 375
Thr Pro Lys Lys Asp Asp Thr Thr Met Thr Lys Ile Ile Gly Asn
380 385 390
Cys Val Ser Ile Leu Val Leu Ser Ser Ala Leu Pro Val Met Ser
395 400 405
Arg Thr Leu Gly Ile Thr Arg Phe Asp Leu Leu Gly Asp Phe Gly
410 415 420
Arg Phe Asn Trp Leu Gly Asn Phe Tyr Ile Val Leu Ser Tyr Asn
425 430 435
Leu Leu Phe Ala Ile Val Thr Thr Leu Cys Leu Val Arg Lys Phe
440 445 450
Thr Ser Ala Val Arg Glu Glu Leu Phe Lys Ala Leu Gly Leu His
455 460 465
Lys Leu His Leu Pro Asn Thr Ser Arg Asp Ser Glu Thr Ala Lys
470 475 480
Pro Ser Val Asn Gly His Gln Lys Ala Leu
485 490

<210> 9
<211> 544
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6013113CD1

<400> 9
Met Met Lys Thr Leu Leu Leu Phe Val Gly Leu Leu Leu Thr Trp
1 5 10 15
Glu Ser Gly Gln Val Leu Gly Asp Gln Thr Val Ser Asp Asn Glu
20 25 30
Leu Gln Glu Met Ser Asn Gln Gly Ser Lys Tyr Val Asn Lys Glu
35 40 45
Ile Gln Asn Ala Val Asn Gly Val Lys Gln Ile Lys Thr Leu Ile
50 55 60
Glu Lys Thr Asn Glu Glu Arg Lys Thr Leu Leu Ser Asn Leu Glu
65 70 75
Glu Ala Lys Lys Lys Lys Glu Asp Ala Leu Asn Glu Thr Arg Glu
80 85 90
Ser Glu Thr Lys Leu Lys Glu Leu Pro Gly Val Cys Asn Glu Thr
95 100 105
Met Met Ala Leu Trp Glu Glu Cys Lys Pro Cys Leu Lys Gln Thr
110 115 120

```
Cys Met Lys Phe Tyr Ala Arg Val Cys Arg Ser Gly Ser Gly Leu
                125                 130                 135
Val Gly Arg Gln Leu Glu Glu Phe Leu Asn Gln Ser Ser Pro Phe
                140                 145                 150
Tyr Phe Trp Met Asn Gly Asp Arg Ile Asp Ser Leu Leu Glu Asn
                155                 160                 165
Asp Arg Gln Gln Thr His Met Leu Asp Val Met Gln Asp His Phe
                170                 175                 180
Ser Arg Ala Ser Ser Ile Ile Asp Glu Leu Phe Gln Asp Arg Phe
                185                 190                 195
Phe Thr Arg Glu Pro Gln Asp Thr Tyr His Tyr Leu Pro Phe Ser
                200                 205                 210
Leu Pro His Arg Arg Pro His Phe Phe Phe Pro Lys Ser Arg Ile
                215                 220                 225
Val Arg Ser Leu Met Pro Phe Ser Pro Tyr Glu Pro Leu Asn Phe
                230                 235                 240
His Ala Met Phe Gln Pro Phe Leu Glu Met Ile His Glu Ala Gln
                245                 250                 255
Gln Ala Met Asp Ile His Phe His Ser Pro Ala Phe Gln His Pro
                260                 265                 270
Pro Thr Glu Phe Ile Arg Glu Gly Asp Asp Asp Arg Thr Val Cys
                275                 280                 285
Arg Glu Ile Arg His Asn Ser Thr Gly Cys Leu Arg Met Lys Asp
                290                 295                 300
Gln Cys Asp Lys Cys Arg Glu Ile Leu Ser Val Asp Cys Ser Thr
                305                 310                 315
Asn Asn Pro Ser Gln Ala Lys Leu Arg Arg Glu Leu Asp Glu Ser
                320                 325                 330
Leu Gln Val Ala Glu Arg Leu Thr Arg Lys Tyr Asn Glu Leu Leu
                335                 340                 345
Lys Ser Tyr Gln Trp Lys Met Leu Asn Thr Ser Ser Leu Leu Glu
                350                 355                 360
Gln Leu Asn Glu Gln Phe Asn Trp Val Ser Arg Leu Ala Asn Leu
                365                 370                 375
Thr Gln Gly Glu Asp Gln Tyr Tyr Leu Arg Val Thr Thr Val Ala
                380                 385                 390
Ser His Thr Ser Asp Ser Asp Val Pro Ser Gly Val Thr Glu Val
                395                 400                 405
Val Val Lys Leu Phe Asp Ser Asp Pro Ile Thr Val Thr Val Pro
                410                 415                 420
Val Glu Val Ser Arg Lys Asn Pro Lys Phe Met Glu Thr Val Ala
                425                 430                 435
Glu Lys Ala Leu Gln Glu Tyr Arg Lys Lys His Arg Asp Ser Leu
                440                 445                 450
Leu Lys Leu Leu Ser Arg Arg Ala Thr Trp Ala Glu Leu Arg Gly
                455                 460                 465
Pro Gly Ala Leu Leu Glu Leu Leu Ala Val Arg Arg Lys Val Ala
                470                 475                 480
Gly Phe Cys Asp Glu Lys Arg Glu Glu Glu Lys Gly Lys Glu Gln
                485                 490                 495
Arg Gly Cys Val Cys Asp Ala Gln Glu Lys Ala Glu Val Ala Val
                500                 505                 510
Lys Leu Leu Arg Asp Glu Gly Gly Arg Ala Leu Cys Asn Cys Gln
                515                 520                 525
Ser Thr Asp Met Gln Gln Gly Pro Phe Leu Ile Val Thr Val Ser
                530                 535                 540
```

Gln Arg Arg Gln

<210> 10
<211> 2758
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488573CD1

<400> 10
Met Asp Val Lys Glu Arg Lys Pro Tyr Arg Ser Leu Thr Arg Arg
1   5   10   15
Arg Asp Ala Glu Arg Arg Tyr Thr Ser Ser Ser Ala Asp Ser Glu
                20   25   30
Glu Gly Lys Ala Pro Gln Lys Ser Tyr Ser Ser Ser Glu Thr Leu
                35   40   45
Lys Ala Tyr Asp Gln Asp Ala Arg Leu Ala Tyr Gly Ser Arg Val
                50   55   60
Lys Asp Ile Val Pro Gln Glu Ala Glu Glu Phe Cys Arg Thr Gly
                65   70   75
Ala Asn Phe Thr Leu Arg Glu Leu Gly Leu Glu Glu Val Thr Pro
                80   85   90
Pro His Gly Thr Leu Tyr Arg Thr Asp Ile Gly Leu Pro His Cys
                95   100   105
Gly Tyr Ser Met Gly Ala Gly Ser Asp Ala Asp Met Glu Ala Asp
                110   115   120
Thr Val Leu Ser Pro Glu His Pro Val Arg Leu Trp Gly Arg Ser
                125   130   135
Thr Arg Ser Gly Arg Ser Ser Cys Leu Ser Ser Arg Ala Asn Ser
                140   145   150
Asn Leu Thr Leu Thr Asp Thr Glu His Glu Asn Thr Glu Thr Pro
                155   160   165
Gly Gly Leu Gln Asn His Ala Arg Leu Arg Thr Pro Pro Pro Pro
                170   175   180
Leu Ser His Ala His Thr Pro Asn Gln His His Ala Ala Ser Ile
                185   190   195
Asn Ser Leu Asn Arg Gly Asn Phe Thr Pro Arg Ser Asn Pro Ser
                200   205   210
Pro Ala Pro Thr Asp His Ser Leu Ser Gly Glu Pro Pro Ala Gly
                215   220   225
Gly Ala Gln Glu Pro Ala His Ala Gln Glu Asn Trp Leu Leu Asn
                230   235   240
Ser Asn Ile Pro Leu Glu Thr Arg Asn Leu Gly Lys Gln Pro Phe
                245   250   255
Leu Gly Thr Leu Gln Asp Asn Leu Ile Glu Met Asp Ile Leu Gly
                260   265   270
Ala Ser Arg His Asp Gly Ala Tyr Ser Asp Gly His Phe Leu Phe
                275   280   285
Lys Pro Gly Gly Thr Ser Pro Leu Phe Cys Thr Thr Ser Pro Gly
                290   295   300
Tyr Pro Leu Thr Ser Ser Thr Val Tyr Ser Pro Pro Pro Arg Pro
                305   310   315
Leu Pro Arg Ser Thr Phe Ala Arg Pro Ala Phe Asn Leu Lys Lys

320 325 330
Pro Ser Lys Tyr Cys Asn Trp Lys Cys Ala Ala Leu Ser Ala Ile
335 340 345
Val Ile Ser Ala Thr Leu Val Ile Leu Leu Ala Tyr Phe Val Gly
350 355 360
Lys His Leu Phe Asn Trp His Leu Gln Pro Met Glu Gly Gln Met
365 370 375
Tyr Glu Ile Thr Glu Asp Thr Ala Ser Ser Trp Pro Val Pro Thr
380 385 390
Asp Val Ser Leu Tyr Pro Ser Gly Gly Thr Gly Leu Glu Thr Pro
395 400 405
Asp Arg Lys Gly Lys Gly Thr Thr Glu Gly Lys Pro Ser Ser Phe
410 415 420
Phe Pro Glu Asp Ser Phe Ile Asp Ser Gly Glu Ile Asp Val Gly
425 430 435
Arg Arg Ala Ser Gln Lys Ile Pro Pro Gly Thr Phe Trp Arg Ser
440 445 450
Gln Val Phe Ile Asp His Pro Val His Leu Lys Phe Asn Val Ser
455 460 465
Leu Gly Lys Ala Ala Leu Val Gly Ile Tyr Gly Arg Lys Gly Leu
470 475 480
Pro Pro Ser His Thr Gln Phe Asp Phe Val Glu Leu Leu Asp Gly
485 490 495
Arg Arg Leu Leu Thr Gln Glu Ala Arg Ser Leu Glu Gly Thr Pro
500 505 510
Arg Gln Ser Arg Gly Thr Val Pro Pro Ser Ser His Glu Thr Gly
515 520 525
Phe Ile Gln Tyr Leu Asp Ser Gly Ile Trp His Leu Ala Phe Tyr
530 535 540
Asn Asp Gly Lys Glu Ser Glu Val Val Ser Phe Leu Thr Thr Ala
545 550 555
Ile Glu Ser Val Asp Asn Cys Pro Ser Asn Cys Tyr Gly Asn Gly
560 565 570
Asp Cys Ile Ser Gly Thr Cys His Cys Phe Leu Gly Phe Leu Gly
575 580 585
Pro Asp Cys Gly Arg Ala Ser Cys Pro Val Leu Cys Ser Gly Asn
590 595 600
Gly Gln Tyr Met Lys Gly Arg Cys Leu Cys His Ser Gly Trp Lys
605 610 615
Gly Ala Glu Cys Asp Val Pro Thr Asn Gln Cys Ile Asp Val Ala
620 625 630
Cys Ser Asn His Gly Thr Cys Ile Met Gly Thr Cys Ile Cys Asn
635 640 645
Pro Gly Tyr Lys Gly Glu Ser Cys Glu Glu Val Asp Cys Met Asp
650 655 660
Pro Thr Cys Ser Gly Arg Gly Val Cys Val Arg Gly Glu Cys His
665 670 675
Cys Ser Val Gly Trp Gly Gly Thr Asn Cys Glu Thr Pro Arg Ala
680 685 690
Thr Cys Leu Asp Gln Cys Ser Gly His Gly Thr Phe Leu Pro Asp
695 700 705
Thr Gly Leu Cys Ser Cys Asp Pro Ser Trp Thr Gly His Asp Cys
710 715 720
Ser Ile Glu Ile Cys Ala Ala Asp Cys Gly Gly His Gly Val Cys
725 730 735
Val Gly Gly Thr Cys Arg Cys Glu Asp Gly Trp Met Gly Ala Ala

740 745 750
Cys Asp Gln Arg Ala Cys His Pro Arg Cys Ala Glu His Gly Thr
755 760 765
Cys Arg Asp Gly Lys Cys Glu Cys Ser Pro Gly Trp Asn Gly Glu
770 775 780
His Cys Thr Ile Ala His Tyr Leu Asp Arg Val Val Lys Glu Gly
785 790 795
Cys Pro Gly Leu Cys Asn Gly Asn Gly Arg Cys Thr Leu Asp Leu
800 805 810
Asn Gly Trp His Cys Val Cys Gln Leu Gly Trp Arg Gly Ala Gly
815 820 825
Cys Asp Thr Ser Met Glu Thr Ala Cys Gly Asp Ser Lys Asp Asn
830 835 840
Asp Gly Asp Gly Leu Val Asp Cys Met Asp Pro Asp Cys Cys Leu
845 850 855
Gln Pro Leu Cys His Ile Asn Pro Leu Cys Leu Gly Ser Pro Asn
860 865 870
Pro Leu Asp Ile Ile Gln Glu Thr Gln Val Pro Val Ser Gln Gln
875 880 885
Asn Leu His Ser Phe Tyr Asp Arg Ile Lys Phe Leu Val Gly Arg
890 895 900
Asp Ser Thr His Ile Ile Pro Gly Glu Asn Pro Phe Asp Gly Gly
905 910 915
His Ala Cys Val Ile Arg Gly Gln Val Met Thr Ser Asp Gly Thr
920 925 930
Pro Leu Val Gly Val Asn Ile Ser Phe Val Asn Asn Pro Leu Phe
935 940 945
Gly Tyr Thr Ile Ser Arg Gln Asp Gly Ser Phe Asp Leu Val Thr
950 955 960
Asn Gly Gly Ile Ser Ile Ile Leu Arg Phe Glu Arg Ala Pro Phe
965 970 975
Ile Thr Gln Glu His Thr Leu Trp Leu Pro Trp Asp Arg Phe Phe
980 985 990
Val Met Glu Thr Ile Ile Met Arg His Glu Glu Asn Glu Ile Pro
995 1000 1005
Ser Cys Asp Leu Ser Asn Phe Ala Arg Pro Asn Pro Val Val Ser
1010 1015 1020
Pro Ser Pro Leu Thr Ser Phe Ala Ser Ser Cys Ala Glu Lys Gly
1025 1030 1035
Pro Ile Val Pro Glu Ile Gln Ala Leu Gln Glu Glu Ile Ser Ile
1040 1045 1050
Ser Gly Cys Lys Met Arg Leu Ser Tyr Leu Ser Ser Arg Thr Pro
1055 1060 1065
Gly Tyr Lys Ser Val Leu Arg Ile Ser Leu Thr His Pro Thr Ile
1070 1075 1080
Pro Phe Asn Leu Met Lys Val His Leu Met Val Ala Val Glu Gly
1085 1090 1095
Arg Leu Phe Arg Lys Trp Phe Ala Ala Ala Pro Asp Leu Ser Tyr
1100 1105 1110
Tyr Phe Ile Trp Asp Lys Thr Asp Val Tyr Asn Gln Lys Val Phe
1115 1120 1125
Gly Leu Ser Glu Ala Phe Val Ser Val Gly Tyr Glu Tyr Glu Ser
1130 1135 1140
Cys Pro Asp Leu Ile Leu Trp Glu Lys Arg Thr Thr Val Leu Gln
1145 1150 1155
Gly Tyr Glu Ile Asp Ala Ser Lys Leu Gly Gly Trp Ser Leu Asp

```
                1160                1165                1170
Lys His His Ala Leu Asn Ile Gln Ser Gly Ile Leu His Lys Gly
                1175                1180                1185
Asn Gly Glu Asn Gln Phe Val Ser Gln Gln Pro Pro Val Ile Gly
                1190                1195                1200
Ser Ile Met Gly Asn Gly Arg Arg Arg Ser Ile Ser Cys Pro Ser
                1205                1210                1215
Cys Asn Gly Leu Ala Asp Gly Asn Lys Leu Leu Ala Pro Val Ala
                1220                1225                1230
Leu Thr Cys Gly Ser Asp Gly Ser Leu Tyr Val Gly Asp Phe Asn
                1235                1240                1245
Tyr Ile Arg Arg Ile Phe Pro Ser Gly Asn Val Thr Asn Ile Leu
                1250                1255                1260
Glu Leu Ser His Ser Pro Ala His Lys Tyr Tyr Leu Ala Thr Asp
                1265                1270                1275
Pro Met Ser Gly Ala Val Phe Leu Ser Asp Ser Asn Ser Arg Arg
                1280                1285                1290
Val Phe Lys Ile Lys Ser Thr Val Val Val Lys Asp Leu Val Lys
                1295                1300                1305
Asn Ser Glu Val Val Ala Gly Thr Gly Asp Gln Cys Leu Pro Phe
                1310                1315                1320
Asp Asp Thr Arg Cys Gly Asp Gly Gly Lys Ala Thr Glu Ala Thr
                1325                1330                1335
Leu Thr Asn Pro Arg Gly Ile Thr Val Asp Lys Phe Gly Leu Ile
                1340                1345                1350
Tyr Phe Val Asp Gly Thr Met Ile Arg Arg Ile Asp Gln Asn Gly
                1355                1360                1365
Ile Ile Ser Thr Leu Leu Gly Ser Asn Asp Leu Thr Ser Ala Arg
                1370                1375                1380
Pro Leu Ser Cys Asp Ser Val Met Asp Ile Ser Gln Val His Leu
                1385                1390                1395
Glu Trp Pro Thr Asp Leu Ala Ile Asn Pro Met Asp Asn Ser Leu
                1400                1405                1410
Tyr Val Leu Asp Asn Asn Val Val Leu Gln Ile Ser Glu Asn His
                1415                1420                1425
Gln Val Arg Ile Val Ala Gly Arg Pro Met His Cys Gln Val Pro
                1430                1435                1440
Gly Ile Asp His Phe Leu Leu Ser Lys Val Ala Ile His Ala Thr
                1445                1450                1455
Leu Glu Ser Ala Thr Ala Leu Ala Val Ser His Asn Gly Val Leu
                1460                1465                1470
Tyr Ile Ala Glu Thr Asp Glu Lys Lys Ile Asn Arg Ile Arg Gln
                1475                1480                1485
Val Thr Thr Ser Gly Glu Ile Ser Leu Val Ala Gly Ala Pro Ser
                1490                1495                1500
Gly Cys Asp Cys Lys Asn Asp Ala Asn Cys Asp Cys Phe Ser Gly
                1505                1510                1515
Asp Asp Gly Tyr Ala Lys Asp Ala Lys Leu Asn Thr Pro Ser Ser
                1520                1525                1530
Leu Ala Val Cys Val Asp Gly Glu Leu Tyr Val Ala Asp Leu Gly
                1535                1540                1545
Asn Ile Arg Ile Arg Phe Ile Arg Lys Asn Lys Pro Phe Leu Asn
                1550                1555                1560
Thr Gln Asn Met Tyr Glu Leu Ser Ser Pro Ile Asp Gln Glu Leu
                1565                1570                1575
Tyr Leu Phe Asp Thr Thr Gly Lys His Leu Tyr Thr Gln Ser Leu
```

1580 1585 1590
Pro Thr Gly Asp Tyr Leu Tyr Asn Phe Thr Tyr Thr Gly Asp Gly
1595 1600 1605
Asp Ile Thr Leu Ile Thr Asp Asn Asn Gly Asn Met Val Asn Val
1610 1615 1620
Arg Arg Asp Ser Thr Gly Met Pro Leu Trp Leu Val Val Pro Asp
1625 1630 1635
Gly Gln Val Tyr Trp Val Thr Met Gly Thr Asn Ser Ala Leu Lys
1640 1645 1650
Ser Val Thr Thr Gln Gly His Glu Leu Ala Met Met Thr Tyr His
1655 1660 1665
Gly Asn Ser Gly Leu Leu Ala Thr Lys Ser Asn Glu Asn Gly Trp
1670 1675 1680
Thr Thr Phe Tyr Glu Tyr Asp Ser Phe Gly Arg Leu Thr Asn Val
1685 1690 1695
Thr Phe Pro Thr Gly Gln Val Ser Ser Phe Arg Ser Asp Thr Asp
1700 1705 1710
Ser Ser Val His Val Gln Val Glu Thr Ser Ser Lys Asp Asp Val
1715 1720 1725
Thr Ile Thr Thr Asn Leu Ser Ala Ser Gly Ala Phe Tyr Thr Leu
1730 1735 1740
Leu Gln Asp Gln Val Arg Asn Ser Tyr Tyr Ile Gly Ala Asp Gly
1745 1750 1755
Ser Leu Arg Leu Leu Leu Ala Asn Gly Met Glu Val Ala Leu Gln
1760 1765 1770
Thr Glu Pro His Leu Leu Ala Gly Thr Val Asn Pro Thr Val Gly
1775 1780 1785
Lys Arg Asn Val Thr Leu Pro Ile Asp Asn Gly Leu Asn Leu Val
1790 1795 1800
Glu Trp Arg Gln Arg Lys Glu Gln Ala Arg Gly Gln Val Thr Val
1805 1810 1815
Phe Gly Arg Arg Leu Arg Val His Asn Arg Asn Leu Leu Ser Leu
1820 1825 1830
Asp Phe Asp Arg Val Thr Arg Thr Glu Lys Ile Tyr Asp Asp His
1835 1840 1845
Arg Lys Phe Thr Leu Arg Ile Leu Tyr Asp Gln Ala Gly Arg Pro
1850 1855 1860
Ser Leu Trp Ser Pro Ser Ser Arg Leu Asn Gly Val Asn Val Thr
1865 1870 1875
Tyr Ser Pro Gly Gly Tyr Ile Ala Gly Ile Gln Arg Gly Ile Met
1880 1885 1890
Ser Glu Arg Met Glu Tyr Asp Gln Ala Gly Arg Ile Thr Ser Arg
1895 1900 1905
Ile Phe Ala Asp Gly Lys Thr Trp Ser Tyr Thr Tyr Leu Glu Lys
1910 1915 1920
Ser Met Val Leu Leu Leu His Ser Gln Arg Gln Tyr Ile Phe Glu
1925 1930 1935
Phe Asp Lys Asn Asp Arg Leu Ser Ser Val Thr Met Pro Asn Val
1940 1945 1950
Ala Arg Gln Thr Leu Glu Thr Ile Arg Ser Val Gly Tyr Tyr Arg
1955 1960 1965
Asn Ile Tyr Gln Pro Pro Glu Gly Asn Ala Ser Val Ile Gln Asp
1970 1975 1980
Phe Thr Glu Asp Gly His Leu Leu His Thr Phe Tyr Leu Gly Thr
1985 1990 1995
Gly Arg Arg Val Ile Tyr Lys Tyr Gly Lys Leu Ser Lys Leu Ala

2000 2005 2010
Glu Thr Leu Tyr Asp Thr Thr Lys Val Ser Phe Thr Tyr Asp Glu
2015 2020 2025
Thr Ala Gly Met Leu Lys Thr Ile Asn Leu Gln Asn Glu Gly Phe
2030 2035 2040
Thr Cys Thr Ile Arg Tyr Arg Gln Ile Gly Pro Leu Ile Asp Arg
2045 2050 2055
Gln Ile Phe Arg Phe Thr Glu Glu Gly Met Val Asn Ala Arg Phe
2060 2065 2070
Asp Tyr Asn Tyr Asp Asn Ser Phe Arg Val Thr Ser Met Gln Ala
2075 2080 2085
Val Ile Asn Glu Thr Pro Leu Pro Ile Asp Leu Tyr Arg Tyr Asp
2090 2095 2100
Asp Val Ser Gly Lys Thr Glu Gln Phe Gly Lys Phe Gly Val Ile
2105 2110 2115
Tyr Tyr Asp Ile Asn Gln Ile Ile Thr Thr Ala Val Met Thr His
2120 2125 2130
Thr Lys His Phe Asp Ala Tyr Gly Arg Met Lys Glu Val Gln Tyr
2135 2140 2145
Glu Ile Phe Arg Ser Leu Met Tyr Trp Met Thr Val Gln Tyr Asp
2150 2155 2160
Asn Met Gly Arg Val Val Lys Lys Glu Leu Lys Val Gly Pro Tyr
2165 2170 2175
Ala Asn Thr Thr Arg Tyr Ser Tyr Glu Tyr Asp Ala Asp Gly Gln
2180 2185 2190
Leu Gln Thr Val Ser Ile Asn Asp Lys Pro Leu Trp Arg Tyr Ser
2195 2200 2205
Tyr Asp Leu Asn Gly Asn Leu His Leu Leu Ser Pro Gly Asn Ser
2210 2215 2220
Ala Arg Leu Thr Pro Leu Arg Tyr Asp Ile Arg Asp Arg Ile Thr
2225 2230 2235
Arg Leu Gly Asp Val Gln Tyr Lys Met Asp Glu Asp Gly Phe Leu
2240 2245 2250
Arg Gln Arg Gly Gly Asp Ile Phe Glu Tyr Asn Ser Ala Gly Leu
2255 2260 2265
Leu Ile Lys Ala Tyr Asn Arg Ala Gly Ser Trp Ser Val Arg Tyr
2270 2275 2280
Arg Tyr Asp Gly Leu Gly Arg Arg Val Ser Ser Lys Ser Ser His
2285 2290 2295
Ser His His Leu Gln Phe Phe Tyr Ala Asp Leu Thr Asn Pro Thr
2300 2305 2310
Lys Val Thr His Leu Tyr Asn His Ser Ser Ser Glu Ile Thr Ser
2315 2320 2325
Leu Tyr Tyr Asp Leu Gln Gly His Leu Phe Ala Met Glu Leu Ser
2330 2335 2340
Ser Gly Asp Glu Phe Tyr Ile Ala Cys Asp Asn Ile Gly Thr Pro
2345 2350 2355
Leu Ala Val Phe Ser Gly Thr Gly Leu Met Ile Lys Gln Ile Leu
2360 2365 2370
Tyr Thr Ala Tyr Gly Glu Ile Tyr Met Asp Thr Asn Pro Asn Phe
2375 2380 2385
Gln Ile Ile Ile Gly Tyr His Gly Gly Leu Tyr Asp Pro Leu Thr
2390 2395 2400
Lys Leu Val His Met Gly Arg Arg Asp Tyr Asp Val Leu Ala Gly
2405 2410 2415
Arg Trp Thr Ser Pro Asp His Glu Leu Trp Lys His Leu Ser Ser

2420 2425 2430
Ser Asn Val Met Pro Phe Asn Leu Tyr Met Phe Lys Asn Asn Asn
2435 2440 2445
Pro Ile Ser Asn Ser Gln Asp Ile Lys Cys Phe Met Thr Asp Val
2450 2455 2460
Asn Ser Trp Leu Leu Thr Phe Gly Phe Gln Leu His Asn Val Ile
2465 2470 2475
Pro Gly Tyr Pro Lys Pro Asp Met Asp Ala Met Glu Pro Ser Tyr
2480 2485 2490
Glu Leu Ile His Thr Gln Met Lys Thr Gln Glu Trp Asp Asn Ser
2495 2500 2505
Lys Ser Ile Leu Gly Val Gln Cys Glu Val Gln Lys Gln Leu Lys
2510 2515 2520
Ala Phe Val Thr Leu Glu Arg Phe Asp Gln Leu Tyr Gly Ser Thr
2525 2530 2535
Ile Thr Ser Cys Gln Gln Ala Pro Lys Thr Lys Lys Phe Ala Ser
2540 2545 2550
Ser Gly Ser Val Phe Gly Lys Gly Val Lys Phe Ala Leu Lys Asp
2555 2560 2565
Gly Arg Val Thr Thr Asp Ile Ile Ser Val Ala Asn Glu Asp Gly
2570 2575 2580
Arg Arg Val Ala Ala Ile Leu Asn His Ala His Tyr Leu Glu Asn
2585 2590 2595
Leu His Phe Thr Ile Asp Gly Val Asp Thr His Tyr Phe Val Lys
2600 2605 2610
Pro Gly Pro Ser Glu Gly Asp Leu Ala Ile Leu Gly Leu Ser Gly
2615 2620 2625
Gly Arg Arg Thr Leu Glu Asn Gly Val Asn Val Thr Val Ser Gln
2630 2635 2640
Ile Asn Thr Val Leu Asn Gly Arg Thr Arg Arg Tyr Thr Asp Ile
2645 2650 2655
Gln Leu Gln Tyr Gly Ala Leu Cys Leu Asn Thr Arg Tyr Gly Thr
2660 2665 2670
Thr Leu Asp Glu Glu Lys Ala Arg Val Leu Glu Leu Ala Arg Gln
2675 2680 2685
Arg Ala Val Arg Gln Ala Trp Ala Arg Glu Gln Gln Arg Leu Arg
2690 2695 2700
Glu Gly Glu Glu Gly Leu Arg Ala Trp Thr Glu Gly Glu Lys Gln
2705 2710 2715
Gln Val Leu Ser Thr Gly Arg Val Gln Gly Tyr Asp Gly Phe Phe
2720 2725 2730
Val Ile Ser Val Glu Gln Tyr Pro Glu Leu Ser Asp Ser Ala Asn
2735 2740 2745
Asn Ile His Phe Met Arg Gln Ser Glu Met Gly Arg Arg
2750 2755

<210> 11
<211> 1139
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506027CD1

<400> 11

Met Glu Pro Asp Ser Leu Leu Asp Gln Asp Asp Ser Tyr Glu Ser
1 5 10 15
Pro Gln Glu Arg Pro Gly Ser Arg Arg Ser Leu Pro Gly Ser Leu
20 25 30
Ser Glu Lys Ser Pro Ser Met Glu Pro Ser Ala Ala Thr Pro Phe
35 40 45
Arg Val Thr Gly Phe Leu Ser Arg Arg Leu Lys Gly Ser Ile Lys
50 55 60
Arg Thr Lys Ser Gln Pro Lys Leu Asp Arg Asn His Ser Phe Arg
65 70 75
His Ile Leu Pro Gly Phe Arg Ser Ala Ala Ala Ala Ala Ala Asp
80 85 90
Asn Glu Arg Ser His Leu Met Pro Arg Leu Lys Glu Ser Arg Ser
95 100 105
His Glu Ser Leu Leu Ser Pro Ser Ser Ala Val Glu Ala Leu Asp
110 115 120
Leu Ser Met Glu Glu Glu Val Val Ile Lys Pro Val His Ser Ser
125 130 135
Ile Leu Gly Gln Asp Tyr Cys Phe Glu Val Thr Thr Ser Ser Gly
140 145 150
Ser Lys Cys Phe Ser Cys Arg Ser Ala Ala Glu Arg Asp Lys Trp
155 160 165
Met Glu Asn Leu Arg Arg Ala Val His Pro Asn Lys Asp Asn Ser
170 175 180
Arg Arg Val Glu His Ile Leu Lys Leu Trp Val Ile Glu Ala Lys
185 190 195
Asp Leu Pro Ala Lys Lys Lys Tyr Leu Cys Glu Leu Cys Leu Asp
200 205 210
Asp Val Leu Tyr Ala Arg Thr Thr Gly Lys Leu Lys Thr Asp Asn
215 220 225
Val Phe Trp Gly Glu His Phe Glu Phe His Asn Leu Pro Pro Leu
230 235 240
Arg Thr Val Thr Val His Leu Tyr Arg Glu Thr Asp Lys Lys Lys
245 250 255
Lys Lys Glu Arg Asn Ser Tyr Leu Gly Leu Val Ser Leu Pro Ala
260 265 270
Ala Ser Val Ala Gly Arg Gln Phe Val Glu Lys Trp Tyr Pro Val
275 280 285
Val Thr Pro Asn Pro Lys Gly Gly Lys Gly Pro Gly Pro Met Ile
290 295 300
Arg Ile Lys Ala Arg Tyr Gln Thr Ile Thr Ile Leu Pro Met Glu
305 310 315
Met Tyr Lys Glu Phe Ala Glu His Ile Thr Asn His Tyr Leu Gly
320 325 330
Leu Cys Ala Ala Leu Glu Pro Ile Leu Ser Ala Lys Thr Lys Glu
335 340 345
Glu Met Ala Ser Ala Leu Val His Ile Leu Gln Ser Thr Gly Lys
350 355 360
Val Lys Asp Phe Leu Thr Asp Leu Met Met Ser Glu Val Asp Arg
365 370 375
Cys Gly Asp Asn Glu His Leu Ile Phe Arg Glu Asn Thr Leu Ala
380 385 390
Thr Lys Ala Ile Glu Glu Tyr Leu Lys Leu Val Gly Gln Lys Tyr
395 400 405
Leu Gln Asp Ala Leu Gly Glu Phe Ile Lys Ala Leu Tyr Glu Ser
410 415 420

Asp Glu Asn Cys Glu Val Asp Pro Ser Lys Cys Ser Ala Ala Asp
425 430 435
Leu Pro Glu His Gln Gly Asn Leu Lys Met Cys Cys Glu Leu Ala
440 445 450
Phe Cys Lys Ile Ile Asn Ser Tyr Cys Val Phe Pro Arg Glu Leu
455 460 465
Lys Glu Val Phe Ala Ser Trp Arg Gln Glu Cys Ser Ser Arg Gly
470 475 480
Arg Pro Asp Ile Ser Glu Arg Leu Ile Ser Ala Ser Leu Phe Leu
485 490 495
Arg Phe Leu Cys Pro Ala Ile Met Ser Pro Ser Leu Phe Asn Leu
500 505 510
Leu Gln Glu Tyr Pro Asp Asp Arg Thr Ala Arg Thr Leu Thr Leu
515 520 525
Ile Ala Lys Val Thr Gln Asn Leu Ala Asn Phe Ala Lys Phe Gly
530 535 540
Ser Lys Glu Glu Tyr Met Ser Phe Met Asn Gln Phe Leu Glu His
545 550 555
Glu Trp Thr Asn Met Gln Arg Phe Leu Leu Glu Ile Ser Asn Pro
560 565 570
Glu Thr Leu Ser Asn Thr Ala Gly Phe Glu Gly Tyr Ile Asp Leu
575 580 585
Gly Arg Glu Leu Ser Ser Leu His Ser Leu Leu Trp Glu Ala Val
590 595 600
Ser Gln Leu Glu Gln Ser Ile Val Ser Lys Leu Gly Pro Leu Pro
605 610 615
Arg Ile Leu Arg Asp Val His Thr Ala Leu Ser Thr Pro Gly Ser
620 625 630
Gly Gln Leu Pro Gly Thr Asn Asp Leu Ala Ser Thr Pro Gly Ser
635 640 645
Gly Ser Ser Ser Ile Ser Ala Gly Leu Gln Lys Met Val Ile Glu
650 655 660
Asn Asp Leu Ser Gly Ser Ser Gly Val Gln Pro Ser Pro Ala Arg
665 670 675
Ser Ser Ser Tyr Ser Glu Ala Asn Glu Pro Asp Leu Gln Met Ala
680 685 690
Asn Gly Gly Lys Ser Leu Ser Met Val Asp Leu Gln Asp Ala Arg
695 700 705
Thr Leu Asp Gly Glu Ala Gly Ser Pro Ala Gly Pro Asp Val Leu
710 715 720
Pro Thr Asp Gly Gln Ala Ala Ala Ala Gln Leu Val Ala Gly Trp
725 730 735
Pro Ala Arg Ala Thr Pro Val Asn Leu Ala Gly Leu Ala Thr Val
740 745 750
Arg Arg Ala Gly Gln Thr Pro Thr Thr Pro Gly Thr Ser Glu Gly
755 760 765
Ala Pro Gly Arg Pro Gln Leu Leu Ala Pro Leu Ser Phe Gln Asn
770 775 780
Pro Val Tyr Gln Met Ala Ala Gly Leu Pro Leu Ser Pro Arg Gly
785 790 795
Leu Gly Asp Ser Gly Ser Glu Gly His Ser Ser Leu Ser Ser His
800 805 810
Ser Asn Ser Glu Glu Leu Ala Ala Ala Ala Lys Leu Gly Ser Phe
815 820 825
Ser Thr Ala Ala Glu Glu Leu Ala Arg Arg Pro Gly Glu Leu Ala
830 835 840

Arg Arg Gln Met Ser Leu Thr Glu Lys Gly Gly Gln Pro Thr Val
845 850 855
Pro Arg Gln Asn Ser Ala Gly Pro Gln Arg Arg Ile Asp Gln Pro
860 865 870
Pro Pro Pro Pro Pro Pro Pro Pro Pro Ala Pro Arg Gly Arg Thr
875 880 885
Pro Pro Asn Leu Leu Ser Thr Leu Gln Tyr Pro Arg Pro Ser Ser
890 895 900
Gly Thr Leu Ala Ser Ala Ser Pro Asp Trp Val Gly Pro Ser Thr
905 910 915
Arg Leu Arg Gln Gln Ser Ser Ser Ser Lys Gly Asp Ser Pro Glu
920 925 930
Leu Lys Pro Arg Ala Val His Lys Gln Gly Pro Ser Pro Val Ser
935 940 945
Pro Asn Ala Leu Asp Arg Thr Ala Ala Trp Leu Leu Thr Met Asn
950 955 960
Ala Gln Leu Leu Glu Asp Glu Gly Leu Gly Pro Asp Pro Pro His
965 970 975
Arg Asp Arg Leu Arg Ser Lys Asp Glu Leu Ser Gln Ala Glu Lys
980 985 990
Asp Leu Ala Val Leu Gln Asp Lys Leu Arg Ile Ser Thr Lys Lys
995 1000 1005
Leu Glu Glu Tyr Glu Thr Leu Phe Lys Cys Gln Glu Glu Thr Thr
1010 1015 1020
Gln Lys Leu Val Leu Glu Tyr Gln Ala Arg Leu Glu Glu Gly Glu
1025 1030 1035
Glu Arg Leu Arg Arg Gln Gln Glu Asp Lys Asp Ile Gln Met Lys
1040 1045 1050
Gly Ile Ile Ser Arg Leu Met Ser Val Glu Glu Glu Leu Lys Lys
1055 1060 1065
Asp His Ala Glu Met Gln Ala Ala Val Asp Ser Lys Gln Lys Ile
1070 1075 1080
Ile Asp Ala Gln Glu Lys Arg Ile Ala Ser Leu Asp Ala Ala Asn
1085 1090 1095
Ala Arg Leu Met Ser Ala Leu Thr Gln Leu Lys Glu Arg Tyr Ser
1100 1105 1110
Met Gln Ala Arg Asn Gly Ile Ser Pro Thr Asn Pro Thr Lys Leu
1115 1120 1125
Gln Ile Thr Glu Asn Gly Glu Phe Arg Asn Ser Ser Asn Cys
1130 1135

<210> 12
<211> 503
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503618CD1

<400> 12
Met Met Lys Thr Leu Leu Leu Phe Val Gly Leu Leu Leu Thr Trp
1 5 10 15
Glu Ser Gly Gln Val Leu Gly Asp Gln Thr Val Ser Asp Asn Glu
20 25 30
Leu Gln Glu Met Ser Asn Gln Gly Ser Lys Tyr Val Asn Lys Glu

35 40 45
Ile Gln Asn Ala Val Asn Gly Val Lys Gln Ile Lys Thr Leu Ile
50 55 60
Glu Lys Thr Asn Glu Glu Arg Lys Thr Leu Leu Ser Asn Leu Glu
65 70 75
Glu Ala Lys Lys Lys Lys Glu Asp Ala Leu Asn Glu Thr Arg Glu
80 85 90
Ser Glu Thr Lys Leu Lys Glu Leu Pro Gly Val Cys Asn Glu Thr
95 100 105
Met Met Ala Leu Trp Glu Glu Cys Lys Pro Cys Leu Lys Gln Thr
110 115 120
Cys Met Lys Phe Tyr Ala Arg Val Cys Arg Ser Gly Ser Gly Leu
125 130 135
Val Gly Arg Gln Leu Glu Glu Phe Leu Asn Gln Ser Ser Pro Phe
140 145 150
Tyr Phe Trp Met Asn Gly Asp Arg Ile Asp Ser Leu Leu Glu Asn
155 160 165
Asp Arg Gln Gln Thr His Met Leu Asp Val Met Gln Asp His Phe
170 175 180
Ser Arg Ala Ser Ser Ile Ile Asp Glu Leu Phe Ser Pro Tyr Glu
185 190 195
Pro Leu Asn Phe His Ala Met Phe Gln Pro Phe Leu Glu Met Ile
200 205 210
His Glu Ala Gln Gln Ala Met Asp Ile His Phe His Ser Pro Ala
215 220 225
Phe Gln His Pro Pro Thr Glu Phe Ile Arg Glu Gly Asp Asp Asp
230 235 240
Arg Thr Val Cys Arg Glu Ile Arg His Asn Ser Thr Gly Cys Leu
245 250 255
Arg Met Lys Asp Gln Cys Asp Lys Cys Arg Glu Ile Leu Ser Val
260 265 270
Asp Cys Ser Thr Asn Asn Pro Ser Gln Ala Lys Leu Arg Arg Glu
275 280 285
Leu Asp Glu Ser Leu Gln Val Ala Glu Arg Leu Thr Arg Lys Tyr
290 295 300
Asn Glu Leu Leu Lys Ser Tyr Gln Trp Lys Met Leu Asn Thr Ser
305 310 315
Ser Leu Leu Glu Gln Leu Asn Glu Gln Phe Asn Trp Val Ser Arg
320 325 330
Leu Ala Asn Leu Thr Gln Gly Glu Asp Gln Tyr Tyr Leu Arg Val
335 340 345
Thr Thr Val Ala Ser His Thr Ser Asp Ser Asp Val Pro Ser Gly
350 355 360
Val Thr Glu Val Val Val Lys Leu Phe Asp Ser Asp Pro Ile Thr
365 370 375
Val Thr Val Pro Val Glu Val Ser Arg Lys Asn Pro Lys Phe Met
380 385 390
Glu Thr Val Ala Glu Lys Ala Leu Gln Glu Tyr Arg Lys Lys His
395 400 405
Arg Asp Ser Leu Leu Lys Leu Leu Ser Arg Arg Ala Thr Trp Ala
410 415 420
Glu Leu Arg Gly Pro Gly Ala Leu Leu Glu Leu Leu Ala Val Arg
425 430 435
Arg Lys Val Ala Gly Phe Cys Asp Glu Lys Arg Glu Glu Glu Lys
440 445 450
Gly Lys Glu Gln Arg Gly Cys Val Cys Asp Ala Gln Glu Lys Ala

```
                    455                     460                 465
Glu Val Ala Val Lys Leu Leu Arg Asp Glu Gly Gly Arg Ala Leu
                    470                     475                 480
Cys Asn Cys Gln Ser Thr Asp Met Gln Gln Gly Pro Phe Leu Ile
                    485                     490                 495
Val Thr Val Ser Gln Arg Arg Gln
                    500
```

<210> 13
<211> 3971
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1567742CB1

<400> 13

```
gtgggctggg ggctgcggcg gctccggcgc tgtctccccg cacccgaccg ggcgagccgg 60
ctgggccggc ggggtgaggg aaagcagtgg agtcgggagc agaagcgcta gaggcagtgg 120
tcgtggcgcg gcggcggcgg ctcccctgga ggccggggat gtgggagagg cggtggcagc 180
agcggggagg cggctgctgc tggacccggg ggaaactgct ggctgacagg acacccggga 240
gagacgtgag ggagcctgcg tgccacctct cacccctgag tgaagctggg ctcgagaggt 300
cggccctgtg ctccccgggc cgactggcca gcgggcgcgg ggcgggggcg ggaacccggg 360
ctcgggcccg gccgggcgcc gggcggcggc ggccgtggag cagcagcctc ggtgcgacgt 420
ggagggctgg aggcggcggc gatgcactag gcctcgctca gggcggctgc cccgggaccc 480
gcagttgagt ggtgatttta tgcaatggct tcaagccaca gttcttcacc agtgcctcaa 540
ggaagcagca gtgatgtttt ctttaaaata gaggtagatc cgtcaaaaca cattcgacct 600
gtgccatcac tgccagatgt gtgtcccaag gaacccacag gtgattcaca tagtttatat 660
gttgccccat ctctagttac agatcaacat agatggactg tatatcattc caaagtaaat 720
ctcccagcag cattaaacga tcctagatta gcaaaaagag aatctgactt cttcacaaaa 780
acatggggat tggactttgt ggacactgaa gtcatacctt cattctacct cccacagatc 840
agcaaggaac attttacagt atatcaacag gaaatctctc agagagagaa gattcatgag 900
agatgcaaga atatttgtcc tcctaaagat accttcgaaa ggactctttt acatactcat 960
gataaatcca ggacagatct ggagcaagta cctaagattt ttatgaaacc agatttttgcc 1020
ttggatgatt ccttaacttt taattcagtt ttaccatggt ctcattttaa tactgctggt 1080
ggaaaaggaa atcgtgatgc agcttcctca aagttgcttc aagaaaagct gagccattat 1140
ctggatattg tggaagtaaa cattgctcac cagatctctc tacgttcaga agcatttttt 1200
catgcaatga cctctcaaca cgagttgcag gactacctca ggaaaacttc ccaggctgta 1260
aaaatgcttc gagataaaat tgcacagatt gataaagtaa tgtgtgaagg atcactccac 1320
attttaagac tggcacttac cagaaataat tgtgttaaag tatacaataa gctgaagtta 1380
atggccactg tacaccagac tcagcctaca gtacaggtgt tattatctac ttctgaattt 1440
gttggagcat tggacttaat agcaacaaca caagaggttc tacagcagga acttcagggc 1500
attcacagtt tccggcattt gggatcacag ctttgtgaat tagaaaaact gatagataaa 1560
atgatgattg cagaattttc tacttattct cacagtgact taaatagacc actggaagat 1620
gactgtcaag ttttagaaga ggaaagacta atatctcttg tatttggact tttaaaacaa 1680
agaaagctta atttttaga aatctatggt gaaaaaatgg ttattacagc aaagaatatc 1740
attaaacagt gtgtgattaa taaagtttca caaacagaag aaatagacac agatgttgtt 1800
gtgaagcttg cagatcagat gagaatgttg aattttcccc agtggtttga tctgctcaag 1860
gatattttct ctaagtttac aattttccta cagagagtga aggcaacatt aaatatcatt 1920
cacagtgttg ttctctcagt tcttgacaaa aaccaaagga ctagagaatt ggaagagatt 1980
tcacaacaga agaatgctgc aaaagataat tcactggaca cagaggtggc ttatttaatc 2040
catgaaggca tgtttataag tgatgcattc ggtgagggtg agctaacacc tatagcagtt 2100
gacactacct ctcaaagaaa tgcatctcca aatagtgagc cctgcagcag tgattctgta 2160
tccgagccag aatgtactac tgattcttca tccagcaaag agcacacatc atcatctgct 2220
attccaggag gtgtggatat tatggtcagt gaagatatga aattaactga ctcagagcta 2280
```

ggaaagctgg caaataatat ccaggaatta ttatatagtg cctcagatat atgccatgat 2340
cgagctgtca aatttctcat gtcaagagca aaggatggtt ttcttgagaa gctaaattcc 2400
atggaattca taacactttc tagattaatg gaaacattca ttttagacac cgaacagatc 2460
tgtggaagaa aaagcacgtc attacttgga gcacttcaga gccaagctat taagtttgta 2520
aataggtttc atgaagagag aaaaaccaag ctcagcctcc tcttagacaa tgagcgctgg 2580
aagcaagcag atgttcctgc agaatttcag gatcttgttg attctctgtc agatgggaag 2640
attgctttac ctgaaaaaaa atcaggagcc acagaagaaa ggaaaccagc tgaagttctt 2700
attgtcgagg gacaacagta tgcagttgtt ggaaccgtat tgctgttaat aagaattatc 2760
cttgaatatt gccagtgtgt ggataacatc ccatctgtta ctactgacat gcttactcgt 2820
ctgtcagatt tattgaagta cttcaattca agaagttgcc agttagttct tggagctggt 2880
gcactgcaag ttgttggact aaaaacgata actacaaaaa atttggctct ttcttcacga 2940
tgtttgcagt taattgtgca ctacattcct gtgatccggg ctcattttga agctcgacta 3000
ccacctaagc aatatagcat gcttaggcat tttgatcata tcactaagga ctaccatgat 3060
cacatagctg aaatatcagc taagcttgta gcgataatgg atagcttatt tgacaagctg 3120
ttatctaagt atgaagtgaa ggctcctgtt ccttctgcct gtttcaggaa tatttgtaag 3180
caaatgacaa aaatgcacga agctatattt gatctccttc cagaagaaca aacacagatg 3240
ttatttttaa gaattaatgc aagttataaa ctccacttga aaaagcagtt atctcactta 3300
aatgtgataa atgatggagg acctcaaaat gggttggtca cagcagatgt agctttttac 3360
actggaaatc ttcaagcctt aaaaggcctt aaagatttgg acctaaatat ggccgaaatt 3420
tgggagcaga agaggtgatg tcatcctgga aaactgggta gttcatctga ccatgggatg 3480
tgtttgttat gaagaaaatc tggatgcctg tgattcgaga attgaacctg aaacccaaag 3540
tgaactgggg tgggggaagg gaaaaaggaa agtatcaagt gttgggaaac tggattcagt 3600
gggatctaca aggaatgtca tttttgtgca tcctacagtg aggagtaact gatcaggtgt 3660
ctataacatt tttcattctc tctggaaaca gactcaggtt tctttggacc aaatccaaaa 3720
gaacacatag ctgtaacaca gctgtagttg actagaatgc tctgtatact ttatattaaa 3780
aaatgctttg catttcttcc agtgcaatga aattcatatg gtgtcccacc ttatttaatg 3840
atggtacaat ttaaaatctt agtcaacttc tgtagaaagt tttctctatg aaagtaaagc 3900
tgtttgaaaa attattattt ttttacagat ctttctataa aaaataaaca tcttttgatt 3960
gcttggaaaa a 3971

<210> 14
<211> 410
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485501CB1

<400> 14
cccaggagtt ggggatgtcc tacaaaccta ccacccctgc ccccagcagc acccccggct 60
tcagcacccc tgggccaggc actccggtcc ctacaggaag cgtcccgtcg ccgtcgggct 120
cagggccggg agccactgcc ccttgcagac cgctgtttaa agactttgga ccacctacgg 180
tcggttgtgt gcaggccatg aaaccacctg gtgcccaggg ctcccagagc acctacacgg 240
aactgctgtt ggtcacaggg gagatgggca aagggatccg gcccacctat gctggcagca 300
agagcgccgc ggagcgcctg aagagaggta tcatccatcc ctagtcagag tgcctggtag 360
agacagagcg gaacgcccac acttaacagg aagctcctag gcctctgtgt 410

<210> 15
<211> 2597
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3089944CB1

<400> 15
gcgccccgtg agcccgagca cccgggagtc ccgagcctcg cgccccggag tgcccgagcc 60
tgcgccgccg cacccggata ccccgcgtcc ccgcgagctg ccgaggccgc ccgccgccgc 120
cccgcggaca gtaccgcctt cctcccctct gtccgcgcca tggccgcccc cgacctgtcc 180
accaacctcc aggaggaggc cacctgcgcc atctgcctcg actacttcac ggatccggtg 240
atgaccgact gcggccacaa cttctgccgc gagtgcatcc ggcgctgctg gggccagccc 300
gagggcccgt acgcgtgccc cgagtgccgc gagctgtccc cgcagaggaa cctgcggccc 360
aaccgcccgc ttgctaagat ggccgagatg gcgcggcgcc tgcacccgcc gtcgccggtc 420
ccgcagggcg tgtgccccgc gcaccgcgag ccactggccg ccttctgtgg cgacgagctg 480
cgcctcctgt gtgcggcctg cgagcgctct ggggagcact gggcgcaccg cgtgcggccg 540
ctgcaggacg cggccgaaga cctcaaggcg aagctggaga agtcactgga gcatctccgg 600
aagcagatgc aggatgcgtt gctgttccaa gcccaggcgg atgagacctg cgtcttgtgg 660
cagaagatgg tggagagcca gcggcagaac gtgctgggtg agttcgagcg tcttcgccgt 720
ttgctggcag aggaggagca gcagctgctg cagaggctgg aggaggagga gctggaggtg 780
ctgccccggc tgcgggaggg cgcagcccac ctaggccagc agagcgccca cctagctgag 840
ctcatcgccg agctcgaggg ccgctgccag ctgcctgctc tggggctgct gcaggacatc 900
aaggacgccc tgcgcagggt ccaggatgtg aagctgcagc cccagaagt tgtgcctatg 960
gagctgagga ccgtgtgcag ggtcccggga ctggtagaga cactgcggag gtttcgaggg 1020
gacgtgacct tggacccgga caccgccaac cctgagctga tcctgtctga agacaggcgg 1080
agcgtgcagc gggggacct acggcaggcc ctgccggaca gcccagagcg ctttgacccc 1140
ggcccctgcg tgctgggcca ggagcgcttc acctcaggcc gccactactg ggaggtggag 1200
gttggggacc gcaccagctg ggccctgggg gtgtgcaggg agaacgtgaa caggaaggag 1260
aagggcgagc tgtccgcggg caacggcttc tggatcctgg tcttcctggg gagctattac 1320
aattcctcgg aacgggcctt ggctccactc cgggacccac ccaggcgcgt ggggatcttt 1380
ctggactacg aggctggaca tctctctttc tacagtgcca ccgatgggtc actgctattc 1440
atctttcccg agatcccctt ctcggggacg ctgcggcccc tcttctcacc cctgtccagc 1500
agcccgaccc cgatgactat ctgccggccg aaaggtgggt ccggggacac cctggctccc 1560
cagtgactcg ggccctcctg gaggagtcct gttgcctctc ctgcccctcc aggccactga 1620
gtgttttggc cacttggagg acctgggagg agggagtgtg tcctttgagc aagaggagga 1680
actcctggtg cctttctgag cctgcgtggg agaaccccaa ttctagcact ccaggaaact 1740
gtgggagagt gtggggcagg ctccgtcctc cctgggagac ccctccagcc accgggtgcc 1800
acttaatgcc aacagccctt accaaagctg ggagccccat tgccccggca gctctggcct 1860
gtggttccag aagctgagaa aactccactg gggcttgcag aatccagggt tcacctaagc 1920
tgcacagttc ctgcagcttt gccagccccc tgaaagtctt gtgtacccca cctctgaaga 1980
tgctgggggga ggcagctggg atgggagcca gccccatgcc tgtctgtgac cccacagtgg 2040
gtgagagccc gtcacagtcc tgggtgtggc tgctctggaa gaattaggag gcagccataa 2100
taagagtctt cagagagatg atgggagggg ccagtgagga caggaacaga gagtagatgt 2160
cctataataa aggggcttct gggaggtgcc tgggcacaga tgtctgttca gcaggtgtgt 2220
gggcctagag gagagagcag agcccagaaa tgtcttttgc aggcccacgt tctgacttga 2280
agctttcgtg ggcatgttgc cattgggttt tgcccttgca aaggcttcct aggtctccag 2340
tggcccctca ggacccaggg tcccagctgc tgcttgggga tgtgcactgc tggccgccgg 2400
ccttgcagtc tctctaccct ggggaggaac agtggcttct cagagcctgg ggcatacaga 2460
agaaggcagg agttgatttt tgtgttgggt ttggggtttc tttgtcctca aggtactgtt 2520
ctgtttctct ttacccctct gctttattta ttgtaagcat tcccacgtta aataaacttt 2580
ggctgttgtc tacaaaa 2597

<210> 16
<211> 1480
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5284076CB1

<400> 16

ctggaggtct gctcagacga aggtctccat ggcgttagaa gtcttgatgc tcctcgctgt 60
cttgatttgg accggtgctg agaacctcca tgtgaaaata agttgctctc tggactggtt 120
gatggtctca gttatcccag ttgcagaaag cagaaatctg tatatatttg cggatgaatt 180
acatctggga atgggctgcc ctgcaaatcg gatacataca tatgtatatg agtttatata 240
tcttgttcgt gattgtggca tcaggacaag ggtagtttct gaggaaactc tcctttttca 300
aaccgagctg tactttaccc caaggaatat agatcatgac cctcaggaaa tccatttgga 360
gtgttccacc tctaggaaat cagtgtggct tacaccagtt tctactgaga atgaaataaa 420
attggatcct agtcctttta ttgctgactt tcagacaaca gcagaagagt taggattatt 480
atcttctagt ccaaacttgc tctgagctaa aggagaaatg gaaacttgaa gctggtgtta 540
tgtattttgc aggaaaacag tttcattttt tcatagcaaa aatatagttg gtgtatatct 600
ctccttaagt ctctggtttc taaaaaccct acttcagtaa aggtcctgat tagttgatta 660
gtgaatgtgt atttctaaat atttgtattc agtaggggta tggctgatta atttaacatt 720
aactattagg taattcatat tatacattta agttctttct gttctgtgta gaagattcag 780
aaatatgtct tcaaagacaa tgacttgatc taattgataa gaacctccaa taaatatgtt 840
ctaatatttt tcaggaagaa taaagaatag agagagacat ataaatgtgc aagaggcaaa 900
actttgagca tagtgtaaaa tttaacatat taactctcac gaaaggcaaa atccttttat 960
gtgcagatac tttaattcat gtagattttc ctattaatca gtaaagttga atcctaacaa 1020
taatgccatg tgacaaccta tttagattat tccagaatta aattcaattt attttctaga 1080
gctcaagtaa ccactacctt aactgaaatt tgatgttagg tttcccttgt tcctccgaat 1140
ggttcttcca cactcaaaat aattgaatgg ttgagttggt taagcaaaga gttatcctgc 1200
cacctaagag cattcattaa atgattattt attaccacct actttatact atcttccttt 1260
ctttaaacat ggagtctaaa tatgtaatat atcaaaaaat acttctgatt tggtagattt 1320
cttatatcaa gggtgagaat tgaactgtgc cattggctat tcaatagctt attgaatgta 1380
tgttttggat gccacatcct cctggaagca aattttgcca agatactgtt tattattatt 1440
tttaattaaa gtgatactat tccattttca aaaaaaaaaa 1480

<210> 17
<211> 6877
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2899903CB1

<400> 17
gtctcagcct cacctcttag cttttccatc tgcacagccg ggccagatcc ccgcagccag 60
catcacgggc agccaggcca accgtcccgg cgtcttccta ttttagacat ctcgctgcct 120
cagtcccttc taatgtttcc agccaggctg cggggggagg aaaaagaggt tactgctact 180
ttaaatgtac tgtatgaagg cgagggctgg aaaggggcct gcttgcagga atacccagtc 240
atctagttgg aaaagccgcc agatggaata caaaaggagg aacccagacg ctcatggaga 300
cagcctcggt tcataaatca ggtggggcca ggggctgggg gcccacacgc catggagccc 360
gactcccttc tggaccaaga cgactcctac gagtcgcctc aagaaaggcc gggctctcgg 420
cgcagcctgc ctggcagcct ttccgagaag agccccagca tggagccctc ggccgccacg 480
ccgttccggg tcacgggctt cctcagccgc cgcctcaagg gctccatcaa gcgcaccaag 540
agccagccca agctggaccg caaccacagc ttccgccaca tcctgccggg gttccggagc 600
gccgccgccg ccgccgcgga caatgagagg tcccatctga tgccgaggct gaaggagtct 660
cgctcccacg agtccctgct cagccccagc agtgcggtgg aggcgctgga cctcagcatg 720
gaggaagagg tggtcatcaa gcccgtgcac agcagcatcc ttggccagga ctactgcttc 780
gaggtgacga cgtcatcagg aagcaagtgc ttttcctgcc ggtctgcagc tgagcgggat 840
aagtggatgg agaacctccg gcgagcggtg catcccaaca aggacaacag ccggcgtgtg 900
gagcacatcc tgaagctgtg ggtgatcgag gccaaggacc tgccagccaa gaagaagtac 960
ctgtgcgagc tgtgcctgga cgatgtgctc tatgcccgca ccacgggcaa gctcaagacg 1020
gacaatgttt tctggggcga gcacttcgag ttccacaact tgccgcctct gcgcacggtc 1080
actgtccacc tgtaccggga gaccgacaag aagaagaaga aggagcgcaa cagttacctg 1140
ggcctggtga gcctacctgc tgcctcggtg gccgggcggc agttcgtgga gaagtggtac 1200

ccggtggtga cgcccaaccc caagggcggc aagggccctg gacccatgat ccgcatcaag 1260
gcgcgctacc aaaccatcac catcctgccc atggagatgt acaaagagtt cgctgagcac 1320
atcaccaacc actacctggg gctgtgtgca gccctcgagc ccatcctcag tgccaagacc 1380
aaggaggaga tggcatctgc cctggtgcac atcctgcaga gcacgggcaa ggtgaaggac 1440
ttcctgacag acctgatgat gtcagaggtg gaccgctgcg gggacaacga gcacctcatc 1500
ttccgggaga acacactggc caccaaggcc attgaggagt acctcaagct agtgggccag 1560
aagtacctgc aggacgccct aggtgagttc atcaaagcgc tgtatgagtc agatgagaac 1620
tgcgaagtgg atcccagcaa gtgctcggcc gctgacctcc cagagcacca gggcaacctc 1680
aagatgtgct gcgagctggc cttctgcaag atcatcaact cctactgtgt cttcccacgg 1740
gagttgaaag aggtgtttgc ctcgtggagg caggagtgca gcagtcgcgg ccgcccggac 1800
atcagtgagc ggctcatcag cgcctccctc ttcctgcgct tcctctgccc agccatcatg 1860
tcgccctcac tcttcaacct gctgcaggag taccctgatg accgcactgc ccgcaccctc 1920
accctcatcg ccaaggtcac ccagaacctg gccaactttg ccaaatttgg cagcaaggag 1980
gaatacatgt ccttcatgaa ccagttccta gagcatgagt ggaccaacat gcagcgcttc 2040
ctgctggaga tctccaaccc cgagaccctc tccaatacag ccggcttcga gggctacatc 2100
gacctgggcc gcgagctctc cagcctgcac tcactgctct gggaggccgt cagccagctg 2160
gagcagagca tagtatccaa actgggaccc ctgcctcgga tcctgaggga cgtccacaca 2220
gcactgagca ccccaggtag cgggcagctc ccagggacca atgacctggc ctccacaccg 2280
ggctctggca gcagcagcat ctcagctggg ctgcagaaga tggtgattga gaacgatctt 2340
tccggtctga tagatttcac ccggttaccg tctccaaccc ccgaaaacaa ggacttgttt 2400
tttgtcacaa ggtcctccgg ggtccagccc tcacctgccc gcagctcgag ttactcggaa 2460
gccaacgagc ctgatcttca gatggccaac ggtggcaaga gcctctccat ggtggacctc 2520
caggacgccc gcacgctgga tggggaggca ggctccccgg cgggccccga cgtcctcccc 2580
acagatgggc aggccgctgc agctcagctg gtggccgggt ggccggcccg ggcaacccca 2640
gtgaacctgg cagggctggc cacggtgcgg cgggcaggcc agacaccaac cacaccaggc 2700
acctccgagg gcgcgccagg ccggccccag ctgttggcac cgctctcctt ccagaaccct 2760
gtgtaccaga tggcggctgg cctgccgctg tcaccccgtg gccttggcga ctcaggctct 2820
gagggccaca gctccctgag ctcacacagc aacagcgagg agttggcggc tgctgccaag 2880
ctgggaagtt tcagcactgc cgcggaggag ctggctcggc ggcccggtga gctggcacgg 2940
cgacagatgt cactgactga aaaaggcggg cagcccacgg tgccacggca gaacagtgct 3000
ggcccccaga ggaggatcga ccagcctccg cccccacccc cgccgccacc tcctgccccc 3060
cgcggccgga cgccccccaa cctgctgagc accctgcagt acccaagacc ctcaagcgga 3120
accctggcgt cggcctcacc tgattgggtg ggccccagta cccgcctgag gcagcagtcc 3180
tcttcctcca aggggacag cccagaactg aagccacggg cagtgcacaa gcagggccct 3240
tcacctgtga gccccaatgc cctggaccgc acagccgctt ggctcttgac catgaacgcg 3300
cagttgttag aagacgaggg cctgggccca gacccccccc acagggatag gctaaggagt 3360
aaggacgagc tcagccaagc agaaaaggac ctggcggtgc tgcaggacaa gctgcgaatc 3420
tccaccaaga agctggagga gtatgagacc ctgttcaagt gccaggagga gacgacgcag 3480
aagctggtgc tggagtacca ggcacggctg gaggagggcg aggagcggct gcggcggcag 3540
caggaggaca aggacatcca gatgaagggc atcatcagca ggttgatgtc cgtggaggaa 3600
gaactgaaga aggaccacgc agagatgcaa gcggctgtgg actccaaaca gaagatcatt 3660
gatgcccagg agaagcgcat tgcctcgttg gatgccgcca atgcccgcct catgagtgcc 3720
ctgacccagc tgaaagagag gtacagcatg caagcccgta acggcatctc ccccaccaac 3780
cccaccaaat tgcagattac tgagaacggc gagttcagaa acagcagcaa ttgttaacct 3840
gcctgaggag ggaggaagct acccaaggag agggggacta tggtggccaa gggcagggtc 3900
tcggcctggg gaggcaccca cggttgcagc cccagcgcgg gtgtcaggag gccgagcctc 3960
ccctccctgc cgctgtccag gggcggccg cagagggagc caccagagac tgaagcagcg 4020
tgaggcgagg tcgccagccg ctccctgtgg ggtgcgggca gaagagactg cacgctgggg 4080
agtggggaca gcctgatggg gcaggggcc tgccaaaaat atgtctgttg gttcctgaat 4140
gtggtgtgtc cttgtcctcc tggatctggc cgagtgcatg tgtcccccca cacctgtgcc 4200
agggaggggg cttcctggag gggggattca aggctagggg gcctacacct gtggcttccc 4260
ctcgcctcct tggggggccc gggactccct ggcagccagg ccctgtcatg tgggacctgg 4320
cacttggcag atcagttggc aggcaggaag ataggaggac acagagcagg aggtcagtgt 4380
cccctgcctg tctccatccg aagcacctgc cactgcatgc agcctgttgg gaccttcctg 4440
gctgtgagga actgaggatt cctacccacc cacccccctct gaacctgtcc ccagagcacc 4500
acctgctacc ttcttccctg ccttagttgt attgccagat agacccagtg agggccatgg 4560

cttttttcttg tgagctcttg tccctgtggg gaggacccac agcttcccac acctcccaca 4620
caggcccagg ctgatgctct agggctccca gaagccagag atctgggcgg atctggccag 4680
atggctctga gcactgtatc tgccttctcc tggggcccag cacacccagg gcacagtggt 4740
cctgtaggga gtgccacctg gtgctcaccc tgaaagaaaa ggtgatcctt cctctgagtg 4800
atggtttaaa aaaagattct aacgcctgca ggccctgaga aggtggataa ctgtgatttt 4860
ttttcctttc acagtatgca ttagaaacaa aagcccgctt gctcgcttgc tggaacacag 4920
gggcctttta agttgagcgt gcgcactgca tgggaaatag cggccctgga ggatgttaga 4980
cttgctccct ctccaagaca gcagcagcct gcacctgccc cgtgtgtgtg gccggcctcc 5040
tcctcaccct tcccggcccc cggccaagga cccaggcgct gcatacaggg gaggggcgca 5100
ccccacagct ggggccggtt ttcctcagct ctaggctgtt ctgtagctta tctgcccctc 5160
ccccactttc aagacagatg agcaggagct tgggtctctc tcggcccctg tctgttccca 5220
gcccctgcag attctgagca aaggccctgg gtaagaaggg tgggagtggg gcctttgcca 5280
gcagagccag ggcagggcga gctgcaggaa tcacccctct gcccctgcag ctggaatgtg 5340
ccacagaggc cccacctgaa gggtggatgt gctggagggg tggcccagag ccatactgcg 5400
tccaccctga gctcggggac aggtgacagt ggctgctctg ggaaggggct tttagatgta 5460
acctacaatt cagttaggct agagacagat gctggtggag gaagggctgg gccaccaggg 5520
atcacagacc acaggaagat gggaggtgga agcagaggcc ctgcccccac cccttcctgt 5580
ctcactcttc tgtcttgtcc ccacccatgc gccttcgtgc ctgagaccag ggtggccaca 5640
caggcagggc ctggctccag tctcatcctc ccattgccca gtgagccctg ctcttctctc 5700
cccagccccc tcccaccgct gcctcgtaga gtgacctcgg acagagcccc cctagcaata 5760
cagggaggct cccggggcct ggacaggcgg gctcggaggc tacccgctgt ggccggtgcc 5820
agctgccctt gcagggtggg tgagctctca ggccgagagc cttatttacc tagtgcaaaa 5880
actgtaaaag tgtacagact cttcacagat ttttatctta attgcaagtc tgccgatttt 5940
gtaaatgttc ttggtgtttg actgtaatgt aactatctca cctaatggtt gtacatatcc 6000
tttggtcctg gtgctgccga gggctggccg ggactgctgc tctcccaagg gttttatttt 6060
atttctgaat ctagagaaca gtattgggca ggaggaaaag gcttggtgtc tgcgggggt 6120
gtcttccctg cctgtggcat ttgtgtgttg gctttgcagc tgctgtctga gtagtggcca 6180
ctggggtgcc ttcactgggc cagtcaacgg ggggctcctg cccaggccac agagaacctg 6240
agttcccggg agctgggccc tgcctgcagc cagggctggg gttgccagag gccctggagg 6300
gaaggacagt ccctgctggg gaagaacagc cccgggggccc cctggtcacc gagactcagc 6360
ctctgctgga gaaagccacg ccctccctgc tagcacagag gcctgactga ctttttttgct 6420
taacttccat gttctgggtg atggaaactg ccaaacctcc tgtcagtgag gactctttcc 6480
gactgcccag aaagtggggg tggaggaccg aggctacagc tccacacgcc ccggtccccc 6540
agagcatctg ccccaggtac acctcccccct gcgccccgca cgactgcggg agccagactg 6600
tccagggaaa cagcctctct cttttctaca cactcagcca caaagccccc cagctcccac 6660
accgcgtccc agctcccctc ttttgtaagt atgtgaaaag gaaaaaatgc aaacgttgga 6720
gtttgggctg gagctcctcc ctccagctgc gacttttaac tatgtaataa tgtacagagg 6780
aagctgttgg tgttctaaga ctctgtgtgg ctgtgcaatt tctgtacatt tgcaattaga 6840
aatattaaag atttatttag ctattttaaa aaaaaaa 6877

<210> 18
<211> 1290
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491355CB1

<400> 18
atgagccgcg cgcgtggggc gctgtgccgg gcctgcctcg cgctggccgc ggccctggcc 60
gcgctgctgt tactgccgct gccgctgccc cgcgcgcccg ccccggcccg gaccccccgcc 120
ccggccccgc gcgcgccccc gtcccggccc gctgccccca gcctgcggcc tgacgacgtc 180
ttcatcgccg tcaagaccac ccggaagaac cacgggccgc gcctgctgct gctgctgcgc 240
acctggatct cccgggcccg ccagcagacg tttatcttca ccgacgggga cgaccctgag 300
ctcgagctcc agggcggcga ccgtgtcatc aacaccaact gctcggcggt gcgcactcgt 360

caggccctct gctgcaagat gtccgtggag tatgacaagt tcattgagtc cgggcgcaag 420
tggttttgcc acgtggatga tgacaattat gtgaacgcaa ggagcctcct gcacctgctc 480
tccagcttct cacccagcca ggacgtctac ctggggcggc ccagcctgga ccacccccatt 540
gaggccaccg agagggtcca gggtggcaga actgtgacca cggtcaagtt ctggtttgct 600
actggtgggg ccgggttctg cctcagcaga ggccttgccc tcaagatgag cccatggccc 660
agcctgggca gcttcatgag cacagctgag caggtgcggc tgccggatga ctgcacagtt 720
ggctacatcg tggagggggct cctgggcgcc cgcctgctgc acagccccct cttccactct 780
cacctggaga acctgcagag gctgccgccc gacaccctgc tccagcaggt taccttgagc 840
catgggggtc ctgagaaccc acagaacgtg gtgaacgtgg ctggaggctt cagcctgcat 900
caagacccca cacggtttaa gtctatccat tgtcttctgt acccagacac ggactggtgt 960
cccaggcaga aacagggcgc cccgacctct cggtgacacc aaccaccccg acccagggct 1020
gcctggctct gtcccaggcg cggggaacca gagcccccta tgggctcagt ggggctccct 1080
caggtgccac ggccacacca gtgagatgca ggcacctggc agaccctctg gctagcctgc 1140
agcccccccct ctcccagccc ctggtgggtg cggtgatggg tgttttggga gaacgaagac 1200
agccaggctg atggccaggg ccgcagtggc cctccccccg acccagcccc aaggttgatc 1260
tcacgggaac aggcttccac cccagcactc 1290

<210> 19
<211> 2133
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3333288CB1

<400> 19
ctcggcagat gccgcctggt ccagctatcg tgctcggtat tcagttttcc ggagcagcgc 60
tctttctctg gcccgcggag cggtcccgcg gccgagtacc ggattcccga gtttgggagg 120
ctctgctttc ctccttagga cccactttgc cgtcctgggg tggctgcagt tatgtccgcg 180
ctgcgacctc tcctgcttct gctgctgcct ctgtgtcccg gtcctggtcc cggacccggg 240
agcgaggcaa aggtcacccg gagttgtgca gagacccggc aggtgctggg ggcccgggga 300
tatagcttaa acctaatccc tcccgccctg atctcaggtg agcacctccg ggtctgtccc 360
caggagtaca cctgctgttc cagtgagaca gagcagaggc tgatcaggga gactgaggcc 420
accttccgag gcctggtgga ggacagcggc tcctttctgg ttcacacact ggctgccagg 480
cacagaaaat ttgatgagtt ttttctggag atgctctcag tagcccagca ctctctgacc 540
cagctcttct cccactccta cggccgcctg tatgcccagc acgccctcat attcaatggc 600
ctgttctctc ggctgcgaga cttctatggg gaatctggtg aggggttgga tgacaccctg 660
gcggatttct gggcacagct cctggagaga gtgttcccgc tgctgcaccc acagtacagc 720
ttcccccctg actacctgct ctgcctctca cgcttggcct catctaccga tggctctctg 780
cagccctttg ggactcacc ccgccgcctc cgcctgcaga taacccggac cctggtggct 840
gcccgagcct ttgtgcaggg cctggagact ggaagaaatg tggtcagcga agcgcttaag 900
gtgccggtgt ctgaaggctg cagccaggct ctgatgcgtc tcatcggctg tcccctgtgc 960
cggggggtcc cctcacttat gccctgccag ggcttctgcc tcaacgtggt tcgtggctgt 1020
ctcagcagca ggggactgga gcctgactgg ggcaactatc tggatggtct cctgatcctg 1080
gctgataagc tccagggccc cttttccttt gagctgacgg ccgagtccat tggggtgaag 1140
atctcggagg gtttgatgta cctgcaggaa aacagtgcga aggtgtccgc ccaggtgttt 1200
caggagtgcg gcccccccga cccggtgcct gcccgcaacc gtcgagcccc gccgccccgg 1260
gaagaggcgg gccggctgtg gtcgatggtg accgaggagg agcggcccac gacggccgca 1320
ggcaccaacc tgcaccggct ggtgtgggag ctccgcgagc gtctggcccg gatgcggggc 1380
ttctgggccc ggctgtccct gacggtgtgc ggagactctc gcatggcagc ggacgcctcg 1440
ctggaggcgg cgccctgctg gaccggagcc gggcggggcc ggtacttgcc gccagtggtc 1500
gggggctccc cggccgagca ggtcaacaac cccgagctca aggtggacgc ctcgggcccc 1560
gatgtcccga cacggcggcg tcgactacag ctccgggcgg ccacggccag aatgaaaacg 1620
gccgcactgg gacacgacct ggacgggcag gacgcggatg aggatgccag cggctctgga 1680
gggggacagc agtatgcaga tgactggatg gctggggctg tggctccccc agcccggcct 1740

cctcggcctc cataccctcc tagaagggat ggttctgggg gcaaaggagg aggtggcagt 1800
gccccgctaca accagggccg gagcaggagt gggggggcat ctattggttt tcacacccaa 1860
accatcctca ttctctccct ctcagccctg gccctgcttg gacctcgata acgggggact 1920
gagggtgctt gagtaggatg tgagacttca tgggcctggg tcctgttgag ttttttcagt 1980
atcaatttct taaaccaaat tttaaaaaaa acaaggtggg ggggtgctca tctcgtgacc 2040
tctgccaccc acatccttca caaactccat gtttcagtgt ttgagtccat gtttattctg 2100
caaataaatg gtaatgtatt ggaccccctaa aaa 2133

<210> 20
<211> 5162
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488313CB1

<400> 20
tgctcgtctg aggctgctga ggcgacggcc ggtgtcgtgg tcgcggtacc tgttccaaca 60
cggctcgcgg gcccgtgccg gctccggtcc ccggcgcggc tgtccgagcc cctgcggcgg 120
gcggacgatg gtgtggcgga gcacgcggac gcgggcggcg cggcggcggg catgaaggag 180
gatggaaggg caggacgagg tgtcggcgcg ggagcagcac ttccacagcc aagtgcggga 240
gtccacgata tgtttccttc tttttgccat tctctacgtt gtttcctact tcatcatcac 300
aagatacaag agaaaatcag atgaacaaga agatgaagat gccatcgtca acaggatttc 360
gttgttttg agcacgttca ctctcgcagt gtcagctggg gctgttttgc ttttaccctt 420
ctcaatcatc agcaatgaaa tcctgctttc ttttcctcag aactactata ttcagtggct 480
aaatggctcc ctgattcatg gtttgtggaa tcttgcttcc ctttttttcca acctttgttt 540
atttgtattg atgccctttg cctttttctt tctggaatca gaaggctttg ctggcctgaa 600
aaagggaatc cgagcccgca ttttagagac tttggtcatg cttcttcttc ttgcgttact 660
cattcttggg atagtgtggg tagcttcagc actcattgac aacgatgccg caagcatgga 720
atctttatat gatctctggg agttctatct accctattta tattcctgta tatcattgat 780
gggatgtttg ttacttctct tgtgtacacc agttggcctt tctcgtatgt tcacagtgat 840
gggtcagttg ctagtgaagc caacaattct tgaagacctg gatgaacaaa tttatatcat 900
taccttagag gaagaagcac tccagagacg actaaatggg ctgtcttcat cggtggaata 960
caacataatg gagttggaac aagaacttga aaatgtaaag actcttaaga caaaattaga 1020
gaggcgaaaa aaggcttcag catgggaaag aaatttggtg tatcccgctg ttatggttct 1080
ccttcttatt gagacatcca tctcggtcct cttggtggct tgtaatattc tttgcctatt 1140
ggttgatgaa acagcaatgc caaaaggaac aagggggcct ggaataggaa atgcctctct 1200
ttctacgttt ggttttgtgg gagctgcgct tgaaatcatt ttgattttct atcttatggt 1260
gtcctctgtt gtcggcttct atagccttcg atttttttgga aactttactc ccaagaaaga 1320
tgacacaact atgacaaaga tcattggaaa ttgtgtgtcc atcttggttt tgagctctgc 1380
tctgcctgtg atgtcgagaa cactgggaat cactagattt gatctacttg gcgactttgg 1440
aaggtttaat tggctgggaa atttctatat tgtattatcc tacaatttgc tttttgctat 1500
tgtgacaaca ttgtgtctgg tccgaaaatt cacctctgca gttcgagaag aactttttcaa 1560
ggccctaggg cttcataaac ttcacttacc aaatacttca agggattcag aaacagccaa 1620
gccttctgta aatgggcatc agaaagcact gtgagacgca cagacggcgt cttctgccac 1680
caagagaccc gagaactcca gattcacgac attcctgtcc catgtagaag catttccatt 1740
caaccgtggc ccctcttcag aacctagacc tatcagtgcc atttttttttt cataatctac 1800
gaagaacttg gctatggctg atcttttttta aatttaactt tctgatggac cctgtagttt 1860
ccagttaagt gcagattcct tacagacata tagaacagcg cattcttctg tagacatttg 1920
ctcatgttgg taaatacaat cacccatatg aaaaaattgt tttcacctga tatgaaaatg 1980
ttagaaaagg caaactccgg gacttctaaa gatttactta aatcccatta tgtactttat 2040
tcagaatgta gaagctgact tgaaaggcat ccttggtact aagtgaagct tattcagaaa 2100
atgcattttt caaatgcaat ggcaactgct tgtagatatc atttttgcag tgtatgttgg 2160
agctgtaatg gttgcaatta tgtttcttat ttccttaaaa gcaaaaagcg tagtttctga 2220
tttatgttat agaatgatac tgattagact ttgagccaag gggaaaatac taaattcttt 2280

taaacctgga gccttagaga gccacaggaa tatcttctgt tgtacagtct aataagctgt 2340
ggtaggaagt atcatgtaat cacagtttaa tgacagttta tgtatatata taattcagta 2400
ttccctctga taacatagtt gccagtgttt aatacacttg taacttggat ttttacctta 2460
taggctatat gtatactcag ttttttaaag catttttttc agagatcact taattcccca 2520
tgcttctgca atgcatataa aaactataaa tgccgagtgg tagaaactcc tctttcttca 2580
tagtcctcag gctttggtta catttgcata tgccatttga agcctccagc ttttaccagt 2640
ttaacatcca aagttcacag catcagcatt catggtgtaa gaacagtttt gcagtataac 2700
acgatctgat aatcattcag ttattaaatt gtaaataatt attgggatgg tttcttggct 2760
ttaagtccac tgaataaaaa ctatgaaatt gcactctgtg tcaaccatcc actaggatag 2820
aataccgaaa tctgtgcatg caaaaatagg agatgggccc atttgcacac aattcgtagt 2880
tatgcagtct gctatataaa tatgttcaca tgcactgtgt gtatgaaaat agatggtctg 2940
tgttcagaca aaagtaaaac atttttttca aattgttaca tttaaaggtt ttctgggaga 3000
aatttatgaa acgcaggctg tgtctatttg acatcagaaa tttccacttt aaaccaaaat 3060
aataagaaac tttaatctgt atatttacaa cctttgttga gtacacttcc cccttattta 3120
tacgtctgca tttccttccg agcttcacat ctttctaaaa tgcagcttgg ttttaaaata 3180
aaagaacatt cattttgtga ttctaaacaa gcttcagtaa ataccaccag tatagtactg 3240
gtgaatttct cagcataaaa tcgacatacc taaaaagtta ataaaattca gctcttttcc 3300
aatttcattg ttatgcctat tgaagtatta attgccaggt ttgattttta gtgaagcttg 3360
gagtccatac tttgagcaga ccaagtgaag ggaagaacag aaagaaactc aggagtagag 3420
taatatcact tctcacttac accactttca ggcacatcca aagagttcct agatacttgg 3480
aaaatgtctg aaaattttta agtaaaatac taaacttttc agtgtttagc tcaacttttt 3540
gttcatttgg aagtttctct ccatccgagg acttaagcca gttttggatt tgtaagccct 3600
gagtacaata cacttcctgg aggcatcctc actgctgttg aagcaaagga tatgcatggg 3660
gtggaaggac ggcttcgaac ctgggactca tatgccttga gaacaaatag attgttacag 3720
ccttgggctg ctgcgtaatc acggttcctc gaggctcttc ctgagcacat gcccaagcat 3780
ctgcctctgg agagactgac tccaaatgca ggtgcttcca ttggagctag gtcggaggct 3840
gctttatatg acgaactcca gaaatggatg ccagaatacg gaggccaaac gttctgagtc 3900
ctggtaagga cagtcgctct gggggtcctc attttactgc agttcctgca cgccagtgaa 3960
agagaggaga tagaccctgg aaggcagagc tgcagatgct catcatcagg tcaattctgg 4020
agctacagtt ttgtttctga ctggataggg atgcaccagt gactgtcaca tcaagcagtc 4080
cttttattct ctctccttta gtatcgattt taaagggcat taggcactat ggttccagag 4140
tttcttgggg aaaacttgca gattcttatt aattggttct gcaatactta aataaattat 4200
tttacaatta taagttttca gattataaca tttgtattaa tttttactga ttttccaaga 4260
tacttcttag atttactatt tacgtagctt tatgtacatt ctctgtaaaa atagacctct 4320
aaatatgagg ctttacatga aatttgtaca cacatacaca ctaatgttag ctccttaaat 4380
tgctgcacta aggtgctggt tagtagagat ggacggagcc tctcgcgttt tgctctcaga 4440
tgtgttaaag gcgcacgtgt acctgctctc agcggcagtg cggcctcccc atctgctggg 4500
tgcccatggc cctccctgca gcctcagtga tgacctcgtc tgccagggac acaggttttc 4560
atcatttaca ggctcttatg tgctagtttt gttggtagca cgttatttaa tgcataaagg 4620
cagaattctt acaagttttt ttttttaat gtgaacatag atgcagcacc gactttttaa 4680
acttgaaaaa actggtataa tgttaacttt taaaaataac atttggacac actagtaatt 4740
gatttttgtt tacagattgt tttgtttaca aattgttagt ctttgtttct atgagatact 4800
tttagtgtga ctttttaaat gtcttagaaa ttaaaagttg tacaaaaagt gatttcatat 4860
ttggtttata agcatttata tgtggggttt atttgttctt ttgtttttc catcttaaat 4920
atcatcatgg ctaaaactta agggtattta tagtttaatt ccatttcagt tttatagagg 4980
gcagtaatta ttctgatgaa tgttgaatta agaaatggat attttctttc tctgttgtgc 5040
agttattggt agatcaattt cttataaccc acaatgtagc atcaataatt gatagcatgt 5100
atttatttta attacttgaa ttatttagac ttgatttctc taattttttc cataaaagga 5160
ct 5162

<210> 21
<211> 1712
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 6013113CB1

<400> 21
tcccggcttc cagaaagctc cccttgcttt ccgcggcatt ctttgggcag gcgtgcaaag 60
actccagaat tggaggcatg atgaagactc tgctgctgtt tgtggggctg ctgctgacct 120
gggagagtgg gcaggtcctg ggggaccaga cggtctcaga caatgagctc caggaaatgt 180
ccaatcaggg aagtaagtac gtcaataagg aaattcaaaa tgctgtcaac ggggtgaaac 240
agataaagac tctcatagaa aaaacaaacg aagagcgcaa gacactgctc agcaacctag 300
aagaagccaa gaagaagaaa gaggatgccc taaatgagac cagggaatca gagacaaagc 360
tgaaggagct cccaggagtg tgcaatgaga ccatgatggc cctctgggaa gagtgtaagc 420
cctgcctgaa acagacctgc atgaagttct acgcacgcgt ctgcagaagt ggctcaggcc 480
tggttggccg ccagcttgag gagttcctga accagagctc gcccttctac ttctggatga 540
atggtgaccg catcgactcc ctgctggaga acgaccggca gcagacgcac atgctggatg 600
tcatgcagga ccacttcagc cgcgcgtcca gcatcataga cgagctcttc caggacaggt 660
tcttcacccg ggagccccag gatacctacc actacctgcc cttcagcctg ccccaccgga 720
ggcctcactt cttctttccc aagtcccgca tcgtccgcag cttgatgccc ttctctccgt 780
acgagcccct gaacttccac gccatgttcc agcccttcct tgagatgata cacgaggctc 840
agcaggccat ggacatccac ttccacagcc cggccttcca gcacccgcca acagaattca 900
tacgagaagg cgacgatgac cggactgtgt gccgggagat ccgccacaac tccacgggct 960
gcctgcggat gaaggaccag tgtgacaagt gccgggagat cttgtctgtg gactgttcca 1020
ccaacaaccc ctcccaggct aagctgcggc gggagctcga cgaatccctc caggtcgctg 1080
agaggttgac caggaaatac aacgagctgc taaagtccta ccagtggaag atgctcaaca 1140
cctcctcctt gctggagcag ctgaacgagc agtttaactg ggtgtcccgg ctggcaaacc 1200
tcacgcaagg cgaagaccag tactatctgc gggtcaccac ggtggcttcc cacacttctg 1260
actcggacgt tccttccggt gtcactgagg tggtcgtgaa gctctttgac tctgatccca 1320
tcactgtgac ggtccctgta gaagtctcca ggaagaaccc taaatttatg gagaccgtgg 1380
cggagaaagc gctgcaggaa taccgcaaaa agcaccggga cagtttgctg aagctgctaa 1440
gccggagagc cacgtgggct gagctcagag gccctggagc tctcttggag cttctggctg 1500
ttcgccggaa ggtggcagga ttttgtgatg aaaagaggga ggaggagaag ggcaaggagc 1560
aacgagggtg tgtatgtgat gcccaagaga aagcagaggt ggcagtgaag ctcctaagag 1620
acgaaggtgg gagggcactg tgcaactgtc agagcaccga catgcagcag ggtcccttcc 1680
tcatcgtgac tgtcagccag agaaggcagt ga 1712

<210> 22
<211> 8645
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488573CB1

<220>
<221> unsure
<222> 93
<223> a, t, c, g, or other

<400> 22
ggattatttg aaggactatt cttagaccct tttaagaaga tttaaaggaa aaccactcgg 60
ccctgagttc ggcgaggacc ctgtttgtgg atntggagga gcgcgggccg gaggccatgg 120
acgtgaagga gaggaagcct taccgctcgc tgacccggcg ccgcgacgcc gagcgccgct 180
acaccagctc gtccgcggac agcgaggagg gcaaagcccc gcagaaatcg tacagctcca 240
gcgagaccct gaaggcctac gaccaggacg cccgcctagc ctatggcagc cgcgtcaagg 300
acattgtgcc gcaggaggcc gaggaattct gccgcacagg tgccaacttc accctgcggg 360
agctggggct ggaagaagta acgccccctc acgggaccct gtaccggaca gacattggcc 420

tcccccactg cggctactcc atgggggctg gctctgatgc cgacatggag gctgacacgg 480
tgctgtcccc tgagcacccc gtgcgtctgt ggggccggag cacacggtca gggcgcagct 540
cctgcctgtc cagccgggcc aattccaatc tcacactcac cgacaccgag catgaaaaca 600
ctgagactcc gggcggcctg cagaaccacg cgcggctccg gacgccgccg ccgccgctct 660
cgcacgccca cacccccaac cagcaccacg cggcctccat taactccctg aaccggggca 720
acttcacgcc gaggagcaac cccagcccgg cccccacgga ccactcgctc tccggagagc 780
cccctgccgg cggcgcccag gagcctgccc acgcccagga gaactggctg ctcaacagca 840
acatcccccct ggagaccaga aacctaggca agcagccatt cctagggaca ttgcaggaca 900
acctcattga gatggacatt ctcggcgcct cccgccatga tggggcttac agtgacgggc 960
acttcctctt caagcctgga ggcacctccc cgctcttctg caccacatca ccaggtacc 1020
cactgacgtc cagcacagtg tactctcctc cgccccgacc cctgccccgc agcaccttcg 1080
cccggccggc ctttaacctc aagaagccct ccaagtactg taactggaag tgcgcagccc 1140
tgagcgccat cgtcatctca gccactctgg tcatcctgct ggcatacttt gtgggtaagc 1200
acctcttcaa ctggcacctg cagccgatgg aggggcagat gtatgagatc acggaggaca 1260
cagccagcag ttggcctgtg ccaaccgacg tctccctata cccctcaggg ggcactggct 1320
tagagacccc tgacaggaaa ggcaaaggaa ccacagaagg aaagcccagt agtttctttc 1380
cagaggacag tttcatagat tctggagaaa ttgatgtggg aaggcgagct tcccagaaga 1440
ttcctcctgg cactttctgg agatctcaag tgttcataga ccatcctgtg catctgaaat 1500
tcaatgtgtc tctgggaaag gcagccctgg ttggcattta tggcagaaaa ggcctccctc 1560
cttcacatac acagtttgac tttgtggagc tgctggatgg caggaggctc ctaacccagg 1620
aggcgcggag cctagagggg accccgcgcc agtctcgggg aactgtgccc ccctccagcc 1680
atgagacagg cttcatccag tatttggatt caggaatctg gcacttggct ttttacaatg 1740
acggaaagga gtcagaagtg gtttcctttc tcaccactgc cattgagtcg gtggataact 1800
gccccagcaa ctgctatggc aatggtgact gcatctctgg gacctgccac tgcttcctgg 1860
gtttcctggg ccccgactgt ggcagagcct cctgccccgt gctctgtagc ggaaatggcc 1920
aatacatgaa aggcagatgc ttgtgccaca gtggctggaa aggcgctgag tgcgatgtgc 1980
ccaccaacca gtgtatcgat gtggcctgca gcaaccatgg cacctgcatc atgggcacct 2040
gcatctgcaa ccctggctac aagggcgaga gctgtgagga agtggactgc atggacccca 2100
catgttcagg ccggggtgtc tgcgtgagag gcgaatgcca ctgctctgtg ggatggggag 2160
gcaccaactg cgagaccccc agggccacat gcttagacca gtgttcaggc cacggaacct 2220
tcctcccgga caccgggctt tgcagctgtg acccaagctg gactggacac gactgttcta 2280
tcgagatctg tgctgccgac tgtggtggcc atggcgtgtg cgtagggggc acctgccgct 2340
gcgaggatgg ctggatgggg gcagcctgcg accagcgggc ctgccacccg cgctgtgccg 2400
agcatgggac ctgccgcgac ggcaagtgcg agtgcagccc tggctggaat ggcgaacact 2460
gcaccatcgc tcactatctg gatagggtag ttaaagaggg ttgccctggg ttgtgcaatg 2520
gcaacggcag atgtacctta gacctgaatg gttggcactg cgtctgccag ctgggctgga 2580
gaggagctgg ctgtgacact tccatggaga ctgcctgcgg tgacagcaaa gacaatgatg 2640
gagatggcct ggtggactgc atggaccctg actgctgcct ccagcccctg tgccatatca 2700
acccgctgtg ccttggctcc cctaaccctc tggacatcat ccaggagaca caggtccctg 2760
tgtcacagca gaacctacac tccttctatg accgcatcaa gttcctcgtg ggcagggaca 2820
gcacgcacat aatccccggg gagaacccct ttgatggagg gcatgcttgt gttattcgtg 2880
gccaagtgat gacatcagat ggaacccccc tggttggtgt gaacatcagt tttgtcaata 2940
accctctctt tggatataca atcagcaggc aagatggcag ctttgacttg gtgacaaatg 3000
gcggcatctc catcatcctg cggttcgagc gggcaccttt catcacacag gagcacaccc 3060
tgtggctgcc atgggatcgc ttctttgtca tggaaaccat catcatgaga catgaggaga 3120
atgagattcc cagctgtgac ctgagcaatt ttgcccgccc caacccagtc gtctctccat 3180
ccccactgac gtccttcgcc agctcctgtg cagagaaagg ccccattgtg ccggaaattc 3240
aggctttgca ggaggaaatc tctatctctg gctgcaagat gaggctgagc tacctgagca 3300
gccggacccc tggctacaaa tctgtcctga ggatcagcct cacccacccg accatcccct 3360
tcaacctcat gaaggtgcac ctcatggtag cggtggaggg ccgcctcttc aggaagtggt 3420
tcgctgcagc cccagacctg tcctattatt tcatttggga caagacagac gtctacaacc 3480
agaaggtgtt tgggctttca gaagcctttg tttccgtggg ttatgaatat gaatcctgcc 3540
cagatctaat cctgtgggaa aaaagaacaa cagtgctgca gggctatgaa attgatgcgt 3600
ccaagcttgg aggatggagc ctagacaaac atcatgccct caacattcaa agtggcatcc 3660
tgcacaaagg gaatggggag aaccagtttg tgtctcagca gcctcctgtc attgggagca 3720
tcatgggcaa tgggcgccgg agaagcatct cctgccccag ctgcaacggc cttgctgacg 3780

gcaacaagct cctggcccca gtggccctca cctgtggctc tgacgggagc ctctatgtgg 3840
gtgatttcaa ctacattaga aggatcttcc cctctggaaa tgtcaccaac atcctagagc 3900
tgagtcacag tccagcacac aaatactacc tggccacaga ccccatgagt ggggccgtct 3960
tcctttctga cagcaacagc cggcgggtct ttaaaatcaa gtccactgtg gtggtgaagg 4020
accttgtcaa gaactctgag gtggttgcgg ggacaggtga ccagtgcctc ccctttgatg 4080
acactcgctg cggggatggt gggaaggcca cagaagccac actcaccaat cccaggggca 4140
ttacagtgga caagtttggg ctgatctact tcgtggatgg caccatgatc agacgcatcg 4200
atcagaatgg gatcatctcc accctgctcg gctctaatga tctcacatca gcccggccac 4260
tcagctgtga ttctgtcatg gatatttccc aggttcacct ggagtggccc acagacttag 4320
ccatcaaccc aatggacaac tcactttatg tcctcgacaa caatgtggtc ctgcaaatct 4380
ctgaaaacca ccaggtgcgc attgtcgccg ggaggcccat gcactgccag gtccctggca 4440
ttgaccactt cctgctaagc aaggtggcca tccacgcaac cctggagtca gccaccgctt 4500
tggctgtttc acacaatggg gtcctgtata ttgctgagac tgatgagaaa aagatcaacc 4560
gcatcaggca ggtcaccact agtggagaga tctcactcgt tgctggggcc cccagtggct 4620
gtgactgtaa aaatgatgcc aactgtgatt gtttttctgg agacgatggt tatgccaagg 4680
atgcaaagtt aaatacccca tcttccttgg ctgtgtgtgt tgatggggag ctctacgtgg 4740
ccgaccttgg gaacatccga attcggttta tccggaagaa caagcctttc ctcaacaccc 4800
agaacatgta tgagctgtct tcaccaattg accaggagct ctatctgttt gataccaccg 4860
gcaagcacct gtacacccaa agcctgccca caggagacta cctgtacaac ttcacctaca 4920
ctggggacgg cgacatcaca ctcatcacag acaacaatgg caacatggta aatgtccgcc 4980
gagactctac tgggatgccc ctctggctgg tggtcccaga tggccaggtg tactgggtga 5040
ccatgggcac caacagtgca ctcaagagtg tgaccacaca aggacacgag ttggccatga 5100
tgacatacca tggcaattcc ggccttctgg caaccaaaag caatgaaaac ggatggacaa 5160
cattttatga gtacgacagc tttggccgcc tgacaaatgt gaccttccct actggccagg 5220
tgagcagttt ccgaagtgat acagacagtt cagtgcatgt ccaggtagag acctccagca 5280
aggatgatgt caccataacc accaacctgt ctgcctcagg cgccttctac acactgctgc 5340
aagaccaagt ccggaacagc tactacatcg ggccgatgg ctccttgcgg ctgctgctgg 5400
ccaacggcat ggaggtggcg ctgcagactg agccccactt gctggctggc accgtcaacc 5460
ccaccgtggg caagaggaat gtcacgctgc ccatcgacaa cggcctcaac ctggtggagt 5520
ggcgccagcg caaagagcag gctcggggcc aggtcactgt ctttgggcgc cggctgcggg 5580
tgcacaaccg aaatctccta tctctggact ttgatcgcgt aacacgcaca gagaagatct 5640
atgatgacca ccgcaagttc acccttcgga ttctgtacga ccaggcgggg cggcccagcc 5700
tctggtcacc cagcagcagg ctgaatggtg tcaacgtgac atactcccct gggggttaca 5760
ttgctggcat ccagaggggc atcatgtctg aaagaatgga atacgaccag gcgggccgca 5820
tcacatccag gatcttcgct gatgggaaga catggagcta cacatactta gagaagtcca 5880
tggtgctgct actacacagc cagaggcagt atatctttga gttcgacaag aatgaccgcc 5940
tctcttctgt gacgatgccc aacgtggcgc ggcagacact agagaccatc cgctcagtgg 6000
gctactacag aaacatctat cagccccctg agggcaatgc ctcagtcata caggacttca 6060
ctgaggatgg gcacctcctt cacaccttct acctgggcac tggccgcagg gtgatataca 6120
agtatggcaa actgtcaaag ctggcagaga cgctctatga caccaccaag gtcagtttca 6180
cctatgacga gacggcaggc atgctgaaga ccatcaacct acagaatgag ggcttcacct 6240
gcaccatccg ctaccgtcag attgggcccc tgattgaccg acagatcttc cgcttcactg 6300
aggaaggcat ggtcaacgcc cgttttgact acaactatga caacagcttc cgggtgacca 6360
gcatgcaggc tgtgatcaac gagaccccac tgcccattga tctctatcgc tatgatgatg 6420
tgtcaggcaa gacagagcag tttgggaagt ttggtgtcat ttactatgac attaaccaga 6480
tcatcaccac agctgtcatg acccacacca agcattttga tgcatatggc aggatgaagg 6540
aagtgcagta tgagatcttc cgctcgctca tgtactggat gaccgtccag tatgataaca 6600
tggggcgagt agtgaagaag gagctgaagg taggacccta cgccaatacc actcgctact 6660
cctatgagta tgatgctgac ggccagctgc agacagtctc catcaatgac aagccactct 6720
ggcgctacag ctacgacctc aatgggaacc tgcacttact gagccctggg aacagtgcac 6780
ggctcacacc actacggtat gacatccgcg accgcatcac tcggctgggt gacgtgcaat 6840
acaagatgga tgaggatggc ttcctgaggc agcggggcgg tgatatcttt gagtacaact 6900
cagctggcct gctcatcaag gcctacaacc gggctggcag ctggagtgtc aggtaccgct 6960
acgatggcct ggggcggcgc gtgtccagca agagcagcca cagccaccac ctgcagttct 7020
tctatgcaga cctgaccaac cccaccaagg tcacccacct gtacaaccac tccagctctg 7080
agatcacctc cctctactac gacttgcaag gacacctctt tgccatggag ctgagcagtg 7140

gtgatgagtt ttacatagct tgtgacaaca tcgggacccc tcttgctgtc tttagtggaa 7200
caggtttgat gatcaagcaa atcctgtaca cagcctatgg ggagatctac atggatacca 7260
accccaactt tcagatcatc ataggctacc atggtggcct ctatgatcca ctcaccaagc 7320
ttgtccacat gggccggcga gattatgatg tgctggccgg acgctggact agcccagacc 7380
acgagctgtg gaagcacctt agtagcagca acgtcatgcc ttttaatctc tatatgttca 7440
aaaacaacaa ccccatcagc aactcccagg acatcaagtg cttcatgaca gatgttaaca 7500
gctggctgct cacctttgga ttccagctac acaacgtgat ccctggttat cccaaaccag 7560
acatggatgc catggaaccc tcctacgagc tcatccacac acagatgaaa acgcaggagt 7620
gggacaacag caagtctatc ctcggggtac agtgtgaagt acagaagcag ctcaaggcct 7680
ttgtcacctt agaacggttt gaccagctct atggctccac aatcaccagc tgccagcagg 7740
ctccaaagac caagaagttt gcatccagcg gctcagtctt tggcaagggg gtcaagtttg 7800
ccttgaagga tggccgagtg accacagaca tcatcagtgt ggccaatgag gatgggcgaa 7860
gggttgctgc catcttgaac catgcccact acctagagaa cctgcacttc accattgatg 7920
gggtggatac ccattacttt gtgaaaccag gaccttcaga aggtgacctg gccatcctgg 7980
gcctcagtgg ggggcggcga accctggaga atggggtcaa cgtcactgtg tcccagatca 8040
acacagtact taatggcagg actagacgct acacagacat ccagctccag tacggggcac 8100
tgtgcttgaa cacacgctac gggacaacgt tggatgagga gaaggcacgg gtcctggagc 8160
tggcccggca gagagccgtg cgccaagcgt gggcccgcga gcagcagaga ctgcgggaag 8220
gggaggaagg cctgcgggcc tggacagagg gggagaagca gcaggtgctg agcacagggc 8280
gggtgcaagg ctacgacggc tttttcgtga tctctgtcga gcagtaccca gaactgtcag 8340
acagcgccaa caacatccac ttcatgagac agagcgagat gggccggagg tgacagagag 8400
gaccaaggac ttcttgccaa agacagctac tcttttgtgg ccgcatacct gactgtgttg 8460
tacttttaaa aaaatgattt tttaacaagt gcagaaacaa aaagatactg gttgcattgt 8520
aactcatgca acatcctttt ttttagaaaa gaaaaacaca gatttggcct tcgcacattt 8580
tttgcaaaga acagaaggta tttttttctg tagtgtgatc acaatgaaaa ctttattgtc 8640
aaaaa 8645

<210> 23
<211> 6812
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506027CB1

<400> 23
gtctcagcct cacctcttag cttttccatc tgcacagccg ggccagatcc ccgcagccag 60
catcacgggc agccaggcca accgtcccgg cgtcttccta ttttagacat ctcgctgcct 120
cagtcccttc taatgtttcc agccaggctg cgggggggagg aaaaagaggt tactgctact 180
ttaaatgtac tgtatgaagg cgagggctgg aaaggggcct gcttgcagga atacccagtc 240
atctagttgg aaaagccgcc agatggaata caaaaggagg aacccagacg ctcatggaga 300
cagcctcggt tcataaatca ggtggggcca ggggctgggg gcccacacgc catggagccc 360
gactcccttc tggaccaaga cgactcctac gagtcgcctc aagaaaggcc gggctctcgg 420
cgcagcctgc ctggcagcct ttccgagaag agccccagca tggagccctc ggccgccacg 480
ccgttccggg tcacgggctt cctcagccgc cgcctcaagg gctccatcaa gcgcaccaag 540
agccagccca agctggaccg caaccacagc ttccgccaca tcctgccggg gttccggagc 600
gccgccgccg ccgccgcgga caatgagagg tcccatctga tgccgaggct gaaggagtct 660
cgctcccacg agtccctgct cagccccagc agtgcggtgg aggcgctgga cctcagcatg 720
gaggaagagg tggtcatcaa gcccgtgcac agcagcatcc ttggccagga ctactgcttc 780
gaggtgacga cgtcatcagg aagcaagtgc ttttcctgcc ggtctgcagc tgagcgggat 840
aagtggatgg agaacctccg gcgagcggtg catcccaaca aggacaacag ccggcgtgtg 900
gagcacatcc tgaagctgtg ggtgatcgag gccaaggacc tgccagccaa gaagaagtac 960
ctgtgcgagc tgtgcctgga cgatgtgctc tatgcccgca ccacgggcaa gctcaagacg 1020
gacaatgttt tctgggggcga gcacttcgag ttccacaact tgccgcctct gcgcacggtc 1080
actgtccacc tgtaccggga gaccgacaag aagaagaaga aggagcgcaa cagttacctg 1140

ggcctggtga gcctacctgc tgcctcggtg gccgggcggc agttcgtgga gaagtggtac 1200
ccggtggtga cgcccaaccc caagggcggc aagggccctg gacccatgat ccgcatcaag 1260
gcgcgctacc aaaccatcac catcctgccc atggagatgt acaaagagtt cgctgagcac 1320
atcaccaacc actacctggg gctgtgtgca gccctcgagc ccatcctcag tgccaagacc 1380
aaggaggaga tggcatctgc cctggtgcac atcctgcaga gcacgggcaa ggtgaaggac 1440
ttcctgacag acctgatgat gtcagaggtg gaccgctgcg gggacaacga gcacctcatc 1500
ttccgggaga acacactggc caccaaggcc attgaggagt acctcaagct agtgggccag 1560
aagtacctgc aggacgccct aggtgagttc atcaaagcgc tgtatgagtc agatgagaac 1620
tgcgaagtgg atcccagcaa gtgctcggcc gctgacctcc cagagcacca gggcaacctc 1680
aagatgtgct gcgagctggc cttctgcaag atcatcaact cctactgtgt cttcccacgg 1740
gagttgaaag aggtgtttgc ctcgtggagg caggagtgca gcagtcgcgg ccgcccggac 1800
atcagtgagc ggctcatcag cgcctccctc ttcctgcgct tcctctgccc agccatcatg 1860
tcgccctcac tcttcaacct gctgcaggag taccctgatg accgcactgc ccgcaccctc 1920
accctcatcg ccaaggtcac ccagaacctg gccaactttg ccaaatttgg cagcaaggag 1980
gaatacatgt ccttcatgaa ccagttccta gagcatgagt ggaccaacat gcagcgcttc 2040
ctgctggaga tctccaaccc cgagaccctc tccaatacag ccggcttcga gggctacatc 2100
gacctgggcc gcgagctctc cagcctgcac tcactgctct gggaggccgt cagccagctg 2160
gagcagagca tagtatccaa actgggaccc ctgcctcgga tcctgaggga cgtccacaca 2220
gcactgagca ccccaggtag cgggcagctc ccagggacca atgacctggc ctccacaccg 2280
ggctctggca gcagcagcat ctcagctggg ctgcagaaga tggtgattga gaacgatctt 2340
tccgggtcct ccggggtcca gccctcacct gcccgcagct cgagttactc ggaagccaac 2400
gagcctgatc ttcagatggc caacggtggc aagagcctct ccatggtgga cctccaggac 2460
gcccgcacgc tggatgggga ggcaggctcc ccggcgggcc ccgacgtcct ccccacagat 2520
gggcaggccg ctgcagctca gctggtggcc gggtggccgg cccgggcaac cccagtgaac 2580
ctggcagggc tggccacggt gcggcgggca ggccagacac caaccacacc aggcacctcc 2640
gagggcgcgc caggccggcc ccagctgttg gcaccgctct ccttccagaa ccctgtgtac 2700
cagatggcgg ctggcctgcc gctgtcaccc cgtggccttg gcgactcagg ctctgagggc 2760
cacagctccc tgagctcaca cagcaacagc gaggagttgg cggctgctgc caagctggga 2820
agtttcagca ctgccgcgga ggagctggct cggcggcccg gtgagctggc acggcgacag 2880
atgtcactga ctgaaaaagg cgggcagccc acggtgccac ggcagaacag tgctggcccc 2940
cagaggagga tcgaccagcc tccgccccca cccccgccgc cacctcctgc cccccgcggc 3000
cggacgcccc ccaacctgct gagcaccctg cagtacccaa gaccctcaag cggaaccctg 3060
gcgtcggcct cacctgattg ggtgggcccc agtacccgcc tgaggcagca gtcctcttcc 3120
tccaaggggg acagcccaga actgaagcca cgggcagtgc acaagcaggg cccttcacct 3180
gtgagcccca atgccctgga ccgcacagcc gcttggctct tgaccatgaa cgcgcagttg 3240
ttagaagacg agggcctggg cccagacccc ccccacaggg ataggctaag gagtaaggac 3300
gagctcagcc aagcagaaaa ggacctggcg gtgctgcagg acaagctgcg aatctccacc 3360
aagaagctgg aggagtatga gaccctgttc aagtgccagg aggagacgac gcagaagctg 3420
gtgctggagt accaggcacg gctggaggag ggcgaggagc ggctgcggcg gcagcaggag 3480
gacaaggaca tccagatgaa gggcatcatc agcaggttga tgtccgtgga ggaagaactg 3540
aagaaggacc acgcagagat gcaagcggct gtggactcca aacagaagat cattgatgcc 3600
caggagaagc gcattgcctc gttggatgcc gccaatgccc gcctcatgag tgccctgacc 3660
cagctgaaag agaggtacag catgcaagcc cgtaacggca tctcccccac caaccccacc 3720
aaattgcaga ttactgagaa cggcgagttc agaaacagca gcaattgtta acctgcctga 3780
ggagggagga agctacccaa ggagaggggg actatggtgg ccaagggcag ggtctcggcc 3840
tggggaggca cccacggttg cagccccagc gcgggtgtca ggaggccgag cctcccctcc 3900
ctgccgctgt ccaggaggcg gccgcagagg gagccaccag agactgaagc agcgtgaggc 3960
gaggtcacca gccgctccct gtggggtgcg ggcagaagag actgcacgct ggggagtggg 4020
gacagcctga tggggcaggg ggcctgccaa aaatatgtct gttggttcct gaatgtggtg 4080
tgtccttgtc ctcctggatc tggccgagtg catgtgtccc cccacacctg tgccagggag 4140
ggggcttcct ggaggggggga ttcaagggct aggggcctac acctgtggct tcccctcgcc 4200
tccttggggg gcccgggact ccctggcagc caggccctgt catgtgggac ctggcacttg 4260
gcagatcagt tggcaggcag gaagatagga ggacacagag caggaggtca gtgtcccctg 4320
cctgtctcca tccgaagcac ctgccactgc atgcagcctg ttgggacctt cctggctgtg 4380
aggaactgag gattcctacc cacccacccc ctctgaacct gtccccagag caccacctgc 4440
taccttcttc cctgccttag ttgtattgcc agatagaccc agtgagggcc atggcttttt 4500

cttgtgagct cttgtccctg tggggaggac ccacagcttc ccacacctcc cacacaggcc 4560
caggctgatg ctctagggct cccagaagcc agagatctgg gcggatctgg ccagatggct 4620
ctgagcactg tatctgcctt ctcctggggc ccagcacacc cagggcacag tggtcctgta 4680
gggagtgcca cctggtgctc accctgaaag aaaaggtgat ccttcctctg agtgatggtt 4740
taaaaaaaag attctaacgc ctgcaggccc tgagaaggtg gataactgtg attttttttc 4800
ctttcacagt atgcattaga aacaaaagcc cgcttgctcg cttgctggaa cacaggggcc 4860
ttttaagttg agcgtgcgca ctgcatggga aatagcggcc ctggaggatg ttagacttgc 4920
tccctctcca agacagcagc agcctgcacc tgccccgtgt gtgtggccgg cctcctcctc 4980
acccttcccg gcccccggcc aaggacccag gcgctgcata caggggaggg gcgcacccca 5040
cagctggggc cggttttcct cagctctagg ctgttctgta gcttatctgc ccctccccca 5100
ctttcaagac agatgagcag gagcttgggt ctctctcggc ccctgtctgt tcccagcccc 5160
tgcagattct gagcaaaggc cctgggtaag aagggtggga gtggggcctt tgccagcaga 5220
gccagggcag ggcgagctgc aggaatcacc cctctgcccc tgcagctgga atgtgccaca 5280
gaggccccac ctgaagggtg gatgtgctgg aggggtggcc cagagccata ctgcgtccac 5340
cctgagctcg gggacaggtg acagtggctg ctctgggaag gggcttttag atgtaaccta 5400
caattcagtt aggctagaga cagatgctgg tggaggaagg gctgggccac cagggatcac 5460
agaccacagg aagatgggag gtggaagcag aggccctgcc cccacccctt cctgtctcac 5520
tcttctgtct tgtccccacc catgcgcctt cgtgcctgag accagggtgg ccacacaggc 5580
agggcctggc tccagtctca tcctcccatt gcccagtgag ccctgctctt ctctccccag 5640
ccccctccca ccgctgcctc gtagagtgac ctcggacaga gccccctag caatacaggg 5700
aggctcccgg ggcctggaca ggcgggctcg gaggctaccc gctgtggccg gtgccagctg 5760
cccttgcagg gtgggtgagc tctcaggccg agagccttat ttacctagtg caaaaactgt 5820
aaaagtgtac agactcttca cagattttta tcttaattgc aagtctgccg attttgtaaa 5880
tgttcttggt gtttgactgt aatgtaacta tctcacctaa tggttgtaca tatcctttgg 5940
tcctggtgct gccgagggct ggccgggact gctgctctcc caagggtttt attttatttc 6000
tgaatctaga gaacagtatt gggcaggagg aaaaggcttg gtgtctgcgg ggggtgtctt 6060
ccctgcctgt ggcatttgtg tgttggcttt gcagctgctg tctgagtagt ggccactggg 6120
gtgccttcac tgggccagtc aacgggggc tcctgcccag gccacagaga acctgagttc 6180
ccgggagctg ggccctgcct gcagccaggg ctggggttgc cagaggccct ggagggaagg 6240
acagtccctg ctggggaaga acagccccgg ggccccctgg tcaccgagac tcagcctctg 6300
ctggagaaag ccacgccctc cctgctagca cagaggcctg actgactttt ttgcttaact 6360
tccatgttct gggtgatgga aactgccaaa cctcctgtca gtgaggactc tttccgactg 6420
cccagaaagt gggggtggag gaccgaggct acagctccac acgccccggt cccccagagc 6480
atctgcccca ggtacacctc cccctgcgcc ccgcacgact gcgggagcca gactgtccag 6540
ggaaacagcc tctctctttt ctacacactc agccacaaag ccccccagct cccacaccgc 6600
gtcccagctc ccctctttg taagtatgtg aaaaggaaaa aatgcaaacg ttggagtttg 6660
ggctggagct cctccctcca gctgcgactt ttaactatgt aataatgtac agaggaagct 6720
gttggtgttc taagactctg tgtggctgtg caatttctgt acatttgcaa ttagaaatat 6780
taaagattta tttagctatt ttaaaaaaaa aa 6812

<210> 24
<211> 1589
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503618CB1

<400> 24
tcccggcttc cagaaagctc cccttgcttt ccgcggcatt ctttgggcag gcgtgcaaag 60
actccagaat tggaggcatg atgaagactc tgctgctgtt tgtggggctg ctgctgacct 120
gggagagtgg gcaggtcctg ggggaccaga cggtctcaga caatgagctc caggaaatgt 180
ccaatcaggg aagtaagtac gtcaataagg aaattcaaaa tgctgtcaac ggggtgaaac 240
agataaagac tctcatagaa aaaacaaacg aagagcgcaa gacactgctc agcaacctag 300
aagaagccaa gaagaagaaa gaggatgccc taaatgagac cagggaatca gagacaaagc 360

```
tgaaggagct cccaggagtg tgcaatgaga ccatgatggc cctctgggaa gagtgtaagc 420
cctgcctgaa acagacctgc atgaagttct acgcacgcgt ctgcagaagt ggctcaggcc 480
tggttggccg ccagcttgag gagttcctga accagagctc gcccttctac ttctggatga 540
atggtgaccg catcgactcc ctgctggaga acgaccggca gcagacgcac atgctggatg 600
tcatgcagga ccacttcagc cgcgcgtcca gcatcataga cgagctcttc tctccgtacg 660
agcccctgaa cttccacgcc atgttccagc ccttccttga gatgatacac gaggctcagc 720
aggccatgga catccacttc cacagcccgg ccttccagca cccgccaaca gaattcatac 780
gagaaggcga cgatgaccgg actgtgtgcc gggagatccg ccacaactcc acgggctgcc 840
tgcggatgaa ggaccagtgt gacaagtgcc gggagatctt gtctgtggac tgttccacca 900
acaacccctc ccaggctaag ctgcggcggg agctcgacga atccctccag gtcgctgaga 960
ggttgaccag gaaatacaac gagctgctaa agtcctacca gtggaagatg ctcaacacct 1020
cctccttgct ggagcagctg aacgagcagt ttaactgggt gtcccggctg gcaaacctca 1080
cgcaaggcga agaccagtac tatctgcggg tcaccacggt ggcttcccac acttctgact 1140
cggacgttcc ttccggtgtc actgaggtgg tcgtgaagct ctttgactct gatcccatca 1200
ctgtgacggt ccctgtagaa gtctccagga agaaccctaa atttatggag accgtggcgg 1260
agaaagcgct gcaggaatac cgcaaaaagc accgggacag tttgctgaag ctgctaagcc 1320
ggagagccac gtgggctgag ctcagaggcc ctggagctct cttggagctt ctggctgttc 1380
gccggaaggt ggcaggattt tgtgatgaaa agagggagga ggagaagggc aaggagcaac 1440
gagggtgtgt atgtgatgcc caagagaaag cagaggtggc agtgaagctc ctaagagacg 1500
aaggtgggag ggcactgtgc aactgtcaga gcaccgacat gcagcagggt cccttcctca 1560
tcgtgactgt cagccagaga aggcagtga                                  1589
```